New **E**ngland **U**niversity **T**ransportation **C**enter

NE University Transportation Center
77 Massachusetts Avenue, E40-279
Cambridge, MA 02139
Phone: 617-253-0753
Fax: 617-258-7570
web.mit.edu/utc

Principal Investigator:	Archon Fung	**Co-Principal Investigator:**	David Luberoff
Title:	Ford Foundation Professor of Democracy and Citizenship	**Title:**	Executive Director, Rappaport Institute for Greater Boston
University:	Harvard Kennedy School	**University:**	Harvard Kennedy School
Email:	Archon_Fung@Harvard.edu	**Email:**	David_Luberoff@Harvard.edu
Phone:	617-495-9846	**Phone:**	617-495-1346

Final Report

Project Title:

Transit Transparency

Project Number:	Project End Date:	Submission Date:
HVDR22-11	8/31/2012	7/29/2012

Summary Report

Transit Transparency: Effective Disclosure through Open Data

By Francisca M. Rojas

Problem

Public transit agencies have employed intelligent systems for determining schedules and routes and for monitoring the real-time location and status of their vehicle fleets for nearly two decades. But until recently, the data generated by daily operations in the transit system were only available to managers and engineers inside agencies. Transit riders could only consult static information when planning trips, primarily through printed or online timetables or maps. Where dynamic train or bus arrival predictions were accessible, riders could only see this information on fixed signs at transit stations or stops. With the popular adoption of smartphones and other mobile technologies transit riders gained the capacity to access information anywhere and at anytime. Some transit agencies have responded by publicly releasing disaggregated data files for schedules and real-time feeds of vehicle locations. These agencies have thus empowered civic entrepreneurs to innovate in delivering transit information to riders through mobile applications and other technologically-assisted means.

Approach & Methodology

This study examines the process by which transit agencies in the U.S. disclosed their operations data to the public and analyzes how constituencies for that data, particularly software developers and transit riders, used that information. This report is based upon five case histories of public transit agencies – Portland's TriMet, Boston's MBTA, Chicago's CTA, Washington's WMATA, and New York's MTA. We sought to understand the origin, evolution and effect of those agencies' open data initiatives using extensive interview work, web research, and analyses of customer surveys. Through this methodology, we identified the drivers and barriers to adoption of transparent, consumer-oriented information systems by transit agencies.

Findings

Transit agencies adopted transparency strategies in order to create more opportunities for riders to access transit information and therefore improve levels of customer service. Agencies achieved a broader dissemination of transit information through a process of co-production with local software developers who acted as third-party information intermediaries by generating a marketplace of customer-facing digital tools and applications for riders. These local computer programmers reshaped, reproduced and recombined the disaggregated, machine-readable electronic files made available by transit agencies to meet the needs of diverse transit riders, customized to each transit system's unique qualities.

Public disclosure of operations information was a novel approach for transit agencies. They had not anticipated that outside developers — who were also transit riders — would be interested in working with the same raw data that engineers were using internally to manage vast and complex transportation systems. But because agencies had already adopted intelligent transportation systems for internal operations they were able to adapt these technologies to disclose schedule, route and real-time arrival data for public use.

The adaptation process from a closed to an open data strategy was not straightforward. Obstacles to sharing operations data with the public included risk-averse institutional cultures inside transit agencies, proprietary vendor contracts that precluded sharing data with third parties, and time-consuming technical efforts to produce accurate datasets suitable for public disclosure and use.

Open data strategies spread to transit agencies in other U.S. cities due the availability of a data standard for transit schedules and the development of communities of transparency around transit data. Transit agency managers, computer programmers, and transit riders engaged in collaborative efforts to make data accessible, timely, accurate and useful. Together, they co-produced a marketplace of customer-facing transit applications targeting the needs of diverse information users. Transit systems with the greatest number and diversity of third-party applications were those whose agency managers developed strong relationships with local software developers.

At this time, the outcomes of transit transparency efforts are uncertain. There have been few systematic studies that examine whether data disclosure is driving improved performance by transit agencies. In terms of rider effects, preliminary survey results support prior research that mobile access to real-time bus arrival information decreases the perceived and actual wait times for riders. Further, improved access to transit information gives riders greater discretion over their time. Future research will seek to understand whether reduced wait times translate into increased satisfaction with overall transit service over time.

Conclusion

The public disclosure of transit information by agencies is a successful case of open data adoption in the United States. Transit transparency offers insights into the elements that enable effective disclosure and delivery of digital information to the public in cases where there is a strong demand for that information, and where the disclosed information is available at the right place and time for users to act upon.

Recommendations

From the disclosure experiences of the transit agencies examined for this study, we present four recommendations to consider in designing future transparency systems:

1. Identify the problem to be solved with better data.
2. Prioritize the disclosure of data for which there is public demand.
3. Determine whether information intermediaries play a role in the disclosure ecosystem and support the development of that ecosystem.
4. Adopt an open, non-proprietary data standard.

– This page intentionally left blank –

Transit Transparency:
Effective Disclosure through Open Data

By Francisca M. Rojas
Research Director
Transparency Policy Project

Principal Investigators
Archon Fung, Co-director
Transparency Policy Project

David Luberoff, Executive Director
Rappaport Institute for Greater Boston

David Weil, Co-director
Transparency Policy Project

Mary Graham, Co-director
Transparency Policy Project

Transparency Policy Project
Ash Center for Democratic Governance and Innovation
Taubman Center for State and Local Government
Harvard Kennedy School
June 2012

http://transprencypolicy.net
http://ash.harvard.edu
http://www.hks.harvard.edu/centers/taubman

Table of Contents

Acknowledgments

This research was supported by a grant from the U.S. Department of Transportation/New England UTC and internal funding sources from the Ash Center for Democratic Governance and Innovation, the Taubman Center for State and Local Government and the Rappaport Institute for Greater Boston at the Harvard Kennedy School of Government. For their support, I thank Principal investigators Archon Fung, David Weil and Mary Graham, co-directors of the Transparency Policy Project at Harvard University's Ash Center for Democratic Governance and Innovation and David Luberoff, former Executive Director of Harvard's Rappaport Institute for Greater Boston.

This project has also benefitted from significant contributions by Candace Brakewood, PhD student in the Engineering Systems Division at MIT, who was my research partner on the survey work conducted in New York and Boston. Professor Chris Zegras of the Department of Urban Studies and Planning at MIT gave important feedback and guidance on these surveys. I also thank the many public officials and technologists who took the time to share their experiences and insights on transit information disclosure and new technologies.

1. Introduction

Over the past decade, many urban functions have become progressively digitized, with intelligent transportation systems at the forefront of this process of innovation. In the case of public transit, agencies now employ intelligent systems for determining schedules and routes and for monitoring the real-time location and status of their vehicle fleets. Until recently, the vast amounts of data generated by daily operations in the transit system were only available to the managers and engineers inside agencies. Meanwhile, customer-facing information was generally static in nature, primarily available only through printed or online timetables or maps. In the few cases where dynamic, train or bus arrival predictions were accessible to riders, these could only be seen on fixed digital signs located inside transit stations or stops.

The popular adoption of smartphones and other mobile information and communications technologies (ICTs) has shifted how transit riders make decisions about their mobility in the United States' major cities. By adopting an open data strategy and disclosing information on the location of buses and trains, several of the largest transit agencies in the U.S. have empowered civic entrepreneurs to devise innovative ways to deliver dynamic, real-time information to transit riders through mobile phone applications and other technologically-assisted means.

This study examines the process by which some transit agencies in the U.S. disclosed their operations data to the public and analyzes how constituencies for that data, particularly software developers and transit riders, used that information. This report is based upon five case histories of public transit agencies – Portland's TriMet, Boston's MBTA, Chicago's CTA, Washington's WMATA, and New York's MTA (see Appendix III: Case Studies). We sought to understand the origin, evolution and effect of those agencies' open data initiatives using extensive interview work, web research, and analyses of customer surveys. Through this methodology, we identified the drivers and barriers to adoption of transparent, consumer-oriented information systems by transit agencies.

We found that transit agencies' disclosure of operations data improved upon prior customer-information systems because: first, a subset of transit riders with programming skills were able to improve upon existing customer-information systems by customizing schedule, route and real-time arrival data to meet rider needs; second, the development of a data standard for schedule and geospatial information allowed quick adoption by transit agencies and data uptake by independent software developers; and third, the proactive engagement of local software developers by transit agencies fostered a sustainable community of use around transit data.

1.1. Public data releases by transit agencies

At the end of 2005, Portland, Oregon's TriMet became the first transit agency to integrate its schedules and routes with Google Maps. This implied that when Portland residents went to the Google Maps website to get driving directions, they could also see how to take that same trip using public transit. Previously, people could only access transit travel information by visiting the agency's website, by consulting paper schedules and maps, or viewing posted information at transit stops. Adding transit schedules and routes to Google Maps expanded the number of sources TriMet riders could consult to get information about their travel. Although Google is a prime online source for vast amounts of information, this move towards greater disclosure by TriMet was a first step in making transit information more publicly accessible. When TriMet shared its schedules and routes with the online search company, the agency also made a critical decision to release those same digital files to the public. TriMet officials believed that a public-sector agency could not be exclusive with their information to a single entity such as Google. Schedules and routes were a public good: managers and engineers at TriMet felt that their information benefitted the agency and its riders most when it was as widely diffused and accessible as possible.

In the years since, nearly all the major transit agencies in the United States have followed TriMet's lead by expanding the outlets for information related to schedules, routes, service alerts, and real-time locations and arrival predictions for buses and trains. Agencies achieved this by creating public access to the electronic files that encode the geographic and temporal information that are key to riding a transit system. Independent, third-party software developers took these digital files and built a wide variety of customer-facing applications for riders to consult transit information – on mobile phones, computers, even do-it-yourself digital signage.

Integrating schedules and routes into Google Transit drove agencies to reformat internal operations management information for public use. TriMet and Google worked together to develop a data standard for structuring schedule and route information in a non-proprietary and accessible format, which they called the General Transit Feed Specification (GTFS). The GTFS standard presented a model for transit agencies to follow in formatting their data for Google Maps.

Figure 1 depicts the remarkable diffusion of open data systems in transit. Boxes with dashed lines indicate when agencies made available static schedule and route data in the GTFS format for download. As agencies saw that disseminating transit information by electronic means was valuable to their customers, they began to create the points of interaction, or Application Programming Interfaces (APIs), needed for third-party software developers to access dynamic, real-time data feeds of bus and train location information. The solid-line boxes in Figure 1 indicate when agencies made available real-time data feeds to the public. This process gained momentum in 2009 and by the end of 2011 all major transit agencies in the U.S. had 1) posted their routes and schedules on Google Maps, 2) publicly released GTFS files of static transit information, and 3) created APIs for access to real-time information by third-parties.

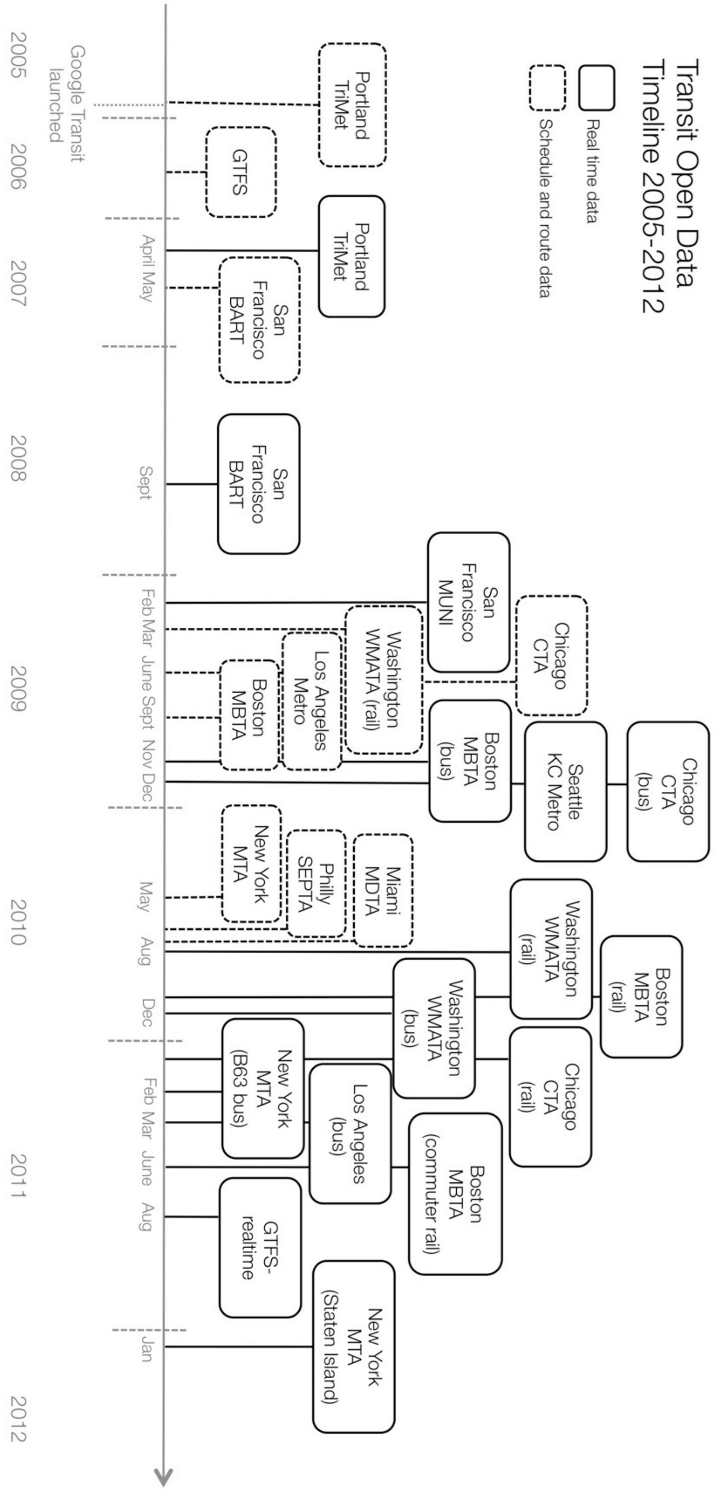


Figure 1 Timetable of transit data releases for largest transit systems in the United States from 2005 to 2012. Dashed-line boxes indicate static route and schedule information. Solid-line boxes indicate real-time data feeds of vehicle locations and arrivals. Source: author's research.

Figure 2 MBTA bus schedules and system map, Harvard Square T station (photo by author)

1.2. Shifting from vertically integrated control to open co-production

Prior to the diffusion of an open data approach for transit data, transit agencies operated vertically-integrated, customer information systems. Agencies controlled all outlets of transit information to the public and managed all of the steps involved in generating that information. Printed schedules and maps were reasonable approaches to disclosing information to transit users in the pre-Internet era, where few other options existed for communicating timetables and routes. Once digital technologies became more widespread, transit agencies posted information about their systems on their websites and by installing electronic signs at stations and stops.

Changing course from customer-information strategies controlled entirely inside agencies to processes where agencies and developers played complementary roles in delivering customized information to riders constituted a significant shift for transit agencies. To varying degrees, factors such as local organizational cultures, technical constraints and political considerations were obstacles that managers in transit agencies had to overcome before being able to disclose disaggregated, machine-readable operations data to the public. Barriers to disclosure of transit data included legal questions as to the ownership of the data, agencies' exposure to risk if third parties were to misrepresent data, or the possibility of embarrassment if disaggregated data were to expose problems in system performance.

Demand for this information from communities of technologists eventually compelled all of the agencies we studied for this report to release schedules, routes and real-time feeds for third-party use.[1] Nevertheless, the obstacles to disclosure mentioned above mediated the level of success achieved by different agencies' open data strategies. The most effective agencies in fostering the co-production of transit information for customers were those that engaged software developers, who then targeted the different needs of particular transit audiences.

1.3. Transit Transparency

According to Fung, Graham and Weil, data disclosure succeeds when it focuses on the specific needs of individuals and groups who are meant to use information to make decisions.[2] Transparency is most effective when disclosing entities provide timely, accurate and complete information when and where people are making important choices, and in standardized formats that everyone can understand. Effective transparency policies also focus on the needs, interests, and capacities of disclosing organizations. They seek to embed new facts in the decision-making routines of information users and to embed user responses into the decision-making of disclosers. The aim is to create new incentives that reduce public risks and improve services.

This analysis of transit data transparency employs Fung, Graham and Weil's "action cycle"[3] framework to trace the process by which software developers integrated information from transit agencies into independent, customer-facing applications for displaying transit data. Figure 3 illustrates the expected action cycle for transit data transparency.

The action cycle is at work when disclosers provide information about their practices to the public, and when that information is useful and accessible for action by target users. The cycle advances when users integrate the disclosed information into their actions or behavior and, in turn, when disclosers adjust their own practices in response. "Transparency policies are effective only when information becomes embedded in this action cycle, becoming an intrinsic part of the decision-making routines of users and disclosers."[4] Thus, the casework for this study traced transit data as it flowed from:

[1] Eric von Hippel's concepts of "user-centered innovation" (2005) and "collective innovation" (2003) help frame the motivation behind technologists' demand for transit data. User-centered innovation is a process whereby certain innovations emerge through product users themselves. In the case of transit, there were certain transit riders with computer programming skills who could improve upon and customize agencies' existing information systems. These programmer/riders shared their work freely on the web through online discussion groups and blogs, spreading the methods they used to "scrape" data and code applications, in line with a "collective innovation" process. These related processes helped compel transit agencies to disclose their operations data in standardized, machine-readable and non-proprietary formats.

[2] This research is summarized in Archon Fung, Mary Graham, and David Weil, *Full Disclosure: The Perils and Performance of Transparency* (2007).

[3] Fung, Graham, and Weil (2007).

[4] Weil, Fung, Graham, and Fagotto (2006).

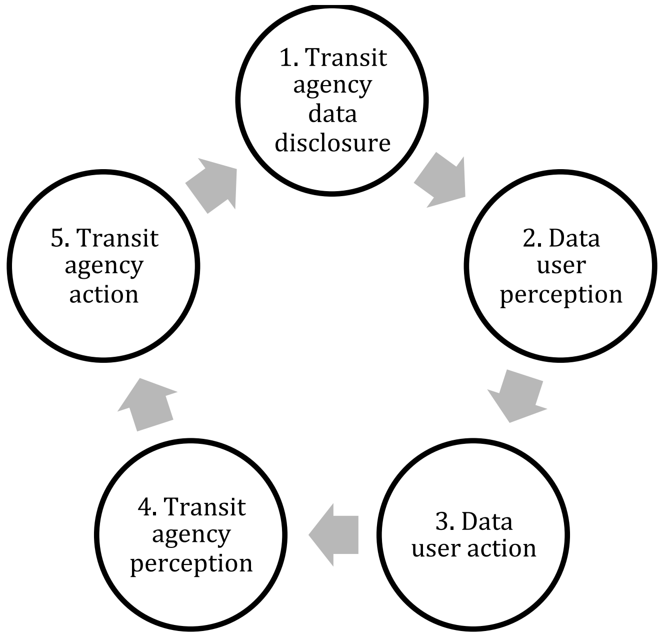


Figure 3 Transparency action cycle (adapted from Fung, Graham and Weil 2007, Pg. 54)

1. the disclosers (transit agencies), to
2. various audiences (developers, transit riders, advocates), to
3. actions taken in response to that information (applications, rider effects, system performance analyses)
4. back to agencies who either did or did not take responsive actions, and
5. outcomes such as improved system performance or increased ridership.

1.4. Research Questions and Case Selection

Based on the transparency action cycle framework, this study of transit data disclosure sought to answer the following questions:

- Why did transit agencies disclose their operations data to the public?
- What conditions enabled the disclosure of data?
- How did the idea to disclose data spread across transit agencies?
- What role did transit agencies, vendors, civic innovators, and riders play in the ecosystem of transparency for transit data?
- Who used transit data and for what purposes?
- What are the outcomes of transit data disclosure?

TABLE 1 TRANSIT AGENCIES BY INNOVATION ADOPTION CATEGORY
(Adapted from Rogers, 1962. Average weekday ridership figures for TriMet, MBTA and CTA from APTA Transit Ridership report for first quarter 2012, representing bus, heavy rail, light rail and trolley ridership. Average weekday ridership figures for WMATA from *Metro Vital Signs Report* for March 2012 and for New York MTA from *MTA Finance Committee Meeting* memo for April 2012.)

Adoption category	Innovator	Early Adopter	Early Majority	Late Majority	Late Adopter
City	Portland	Boston	Chicago	Washington DC	New York
Agency	TriMet	MBTA	CTA	WMATA	MTA
Average weekday ridership	325,500	1,174,900	1,711,900	1,428,900	8,578,575
Year of first data release	2006	2009	2009	2009/2010	2010

To answer the above questions, we selected five cases for in-depth study with guidance from Rogers' diffusion of innovation model, which shows how new ideas and technologies diffuse across organizations.[5] The transit agencies examined for this report correspond with Rogers' five innovation adopter categories (Table 1):[6]

- Innovators: these are the first to adopt an innovation and are willing to take risks. They serve a critical role in influencing others.
- Early adopters: these are the second fastest adopters of an innovation and serve as role models known for "judicious innovation decisions".[7] They help trigger further adoption by others.
- Early majority: these organizations tend to be slower in adopting new ideas because they tend to deliberate for some time before completely jumping on board with an innovation.
- Late majority: these adopt innovations with skepticism and caution, and ultimately adopt and innovation as a result of increasing peer pressure.
- Late adopter: the most tradition-bound, late adopters need to be certain that a new idea will not fail. They tend to engage in lengthy innovation-decision processes.

[5] Rogers, Everett. 2003 (1962). *Diffusion of Innovations*. New York: Free Press.
[6] Rogers (2003), Pg. 281.
[7] Rogers (2003), Pg. 283.

The year of each agency's data release effort determines the corresponding category for each case. Where the dates of agencies' data releases overlap closely, such as in Boston and Chicago, we assigned categories by determining which agency released real-time arrival information first for bus and rail.

1.5. Overview of findings

The five case histories in this study show that transit agencies adopted transparency strategies for their operations data in order to create more opportunities for riders to access transit information and improve levels of customer service. Agencies achieved a broader dissemination of transit information through a process of co-production with local software developers who acted as third-party information intermediaries by generating a marketplace of customer-facing digital tools and applications for riders. These local computer programmers reshaped, reproduced and recombined the disaggregated, machine-readable electronic files made available by transit agencies to meet the needs of diverse transit riders, customized to each transit system's unique qualities.

Public disclosure of operations information was a novel approach for transit agencies. They had not anticipated that outside developers — who were also transit riders — would be interested in working with the same raw data that engineers were using internally to manage such vast and complex transportation systems. Rather, transit agencies had traditionally disseminated information about their systems to customers through agency-designed and -built interfaces accessible at subway stations, bus stops or on agency websites. But because agencies had already adopted intelligent transportation systems for internal operations – for scheduling routes and monitoring the location of buses and trains – they were able to adapt these technologies to disclose schedule, route and real-time arrival data for public use. This process of adaptation from a closed to an open data strategy was not easy for most agencies. Obstacles to sharing operations data with the public included risk-averse institutional cultures inside transit agencies, proprietary vendor contracts that precluded sharing data with third parties, and time-consuming technical efforts to produce accurate datasets suitable for public disclosure and use.

Nevertheless, the open data strategy championed by Portland's TriMet agency spread to agencies in other metropolitan areas due to the communities of transparency that developed around transit data. These communities engaged in collaborative efforts to make data accessible, timely, accurate and useful to transit riders. Transit agency mangers worked behind the scenes to format and release operations data and share their experiences with managers working in other transit systems. Technologists built applications, improved the data, and shared their expertise with other software developers. Transit riders reported software bugs and contributed ideas for new features for transit applications. Together, they co-produced a marketplace of customer-facing transit applications targeting the needs of diverse information users.

Transit systems with the greatest number and diversity of third-party applications were those where key staff developed strong relationships with local software developers. This is the case in Portland, Boston and Chicago. In Washington DC, where WMATA lacks a developer relations mechanism, we found substantially fewer third-party applications built with transit data. New York City, with a moderate commitment to developer relations, has the highest absolute number of third-party applications in part due to the presence of several citywide contests to spur innovation with open data and, as the most extensive transit system in the United States, a large market demand for transit information.

At this time, many of the outcomes of transit transparency efforts are uncertain. There have been few systematic studies that examine whether data disclosure is driving improved performance by transit agencies. But real time data may nevertheless be influencing agency performance and rider response to transit service. In Chicago, for example, real-time bus data was released only after the agency improved on-time arrival metrics. In terms of rider effects, our preliminary survey results support prior research that mobile access to real-time bus arrival information decreases the perceived and actual wait times for riders. Further, improved access to transit information gives riders greater discretion over their time. Our future research will seek to understand whether reduced wait times translate into increased satisfaction with overall transit service over time.

As one of the earliest and arguably most successful cases of open data adoption in the U.S., transit transparency presents a rich example of the elements that enable effective public disclosure and delivery of digital information.

2. The Struggle for Open Data in Transit

2.1. Transit information

Prior to the release of transit information in disaggregated digital formats for third-party use, agencies disseminated information about the functioning of transit systems through means produced entirely within their organizations. The most venerable of these was the printed, paper schedule that provided transit users with the time and location of scheduled bus and subway service throughout the system. The schedules were typically updated on a periodic basis to account for seasonal variation (e.g. winter schedules requiring greater transit time between stops) or changes in service arising from operational decisions (e.g. reducing the number of buses on a route).

Starting in the 1990s, media employed by agencies to provide customer information evolved from:

- paper schedules and maps available at stations (Figures 2 and 4),
- to the same schedules and maps viewable on agency websites,
- to digital signage inside stations and stops displaying arrival information for the next bus or train (Figure 5)
- and online trip planners providing directions from a point of origin to a desired destination (Figure 6).

These traveler information strategies coexisted, thereby expanding the sources of information available to transit riders. Supporting the production of these information sources were specialized technologies – software and hardware alike – that helped transit agencies plan schedules and routes and manage networks of trains and buses.

The deployment by agencies of intelligent transportation systems (ITS) in the late 1980s and throughout the 1990s was initially focused on increasing operational efficiencies by improving the management and control of system operations.[8] Agencies understood the potential customer benefits of these systems only as the technologies matured. Software systems like Hastus and Trapeze facilitated the creation of vehicle and crew schedules and allowed agencies to geographically locate and analyze bus routes and stops. From this same system, marketing departments could print out schedules and maps. Customer-facing information generated by scheduling and routing software constituted a form of static information disclosure. But schedules and routes represented the intentions of a system: where riders should expect a bus or train to be at a given place and time.

[8] Transit Cooperative Research Program (2003).

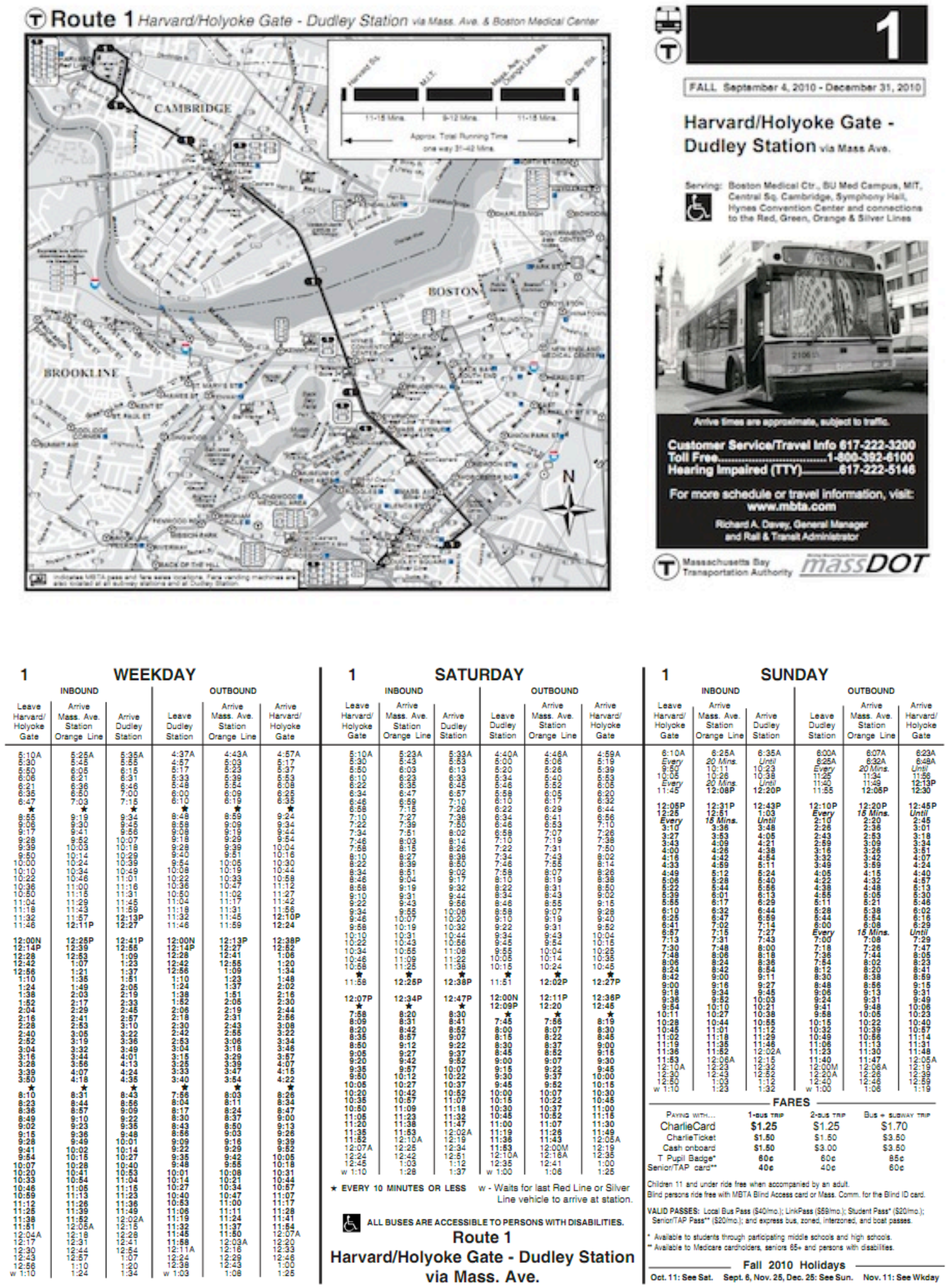


Figure 4 Paper route and schedule for Boston's MBTA #1 bus

Figure 5 Electronic signs at one of Boston's MBTA stations

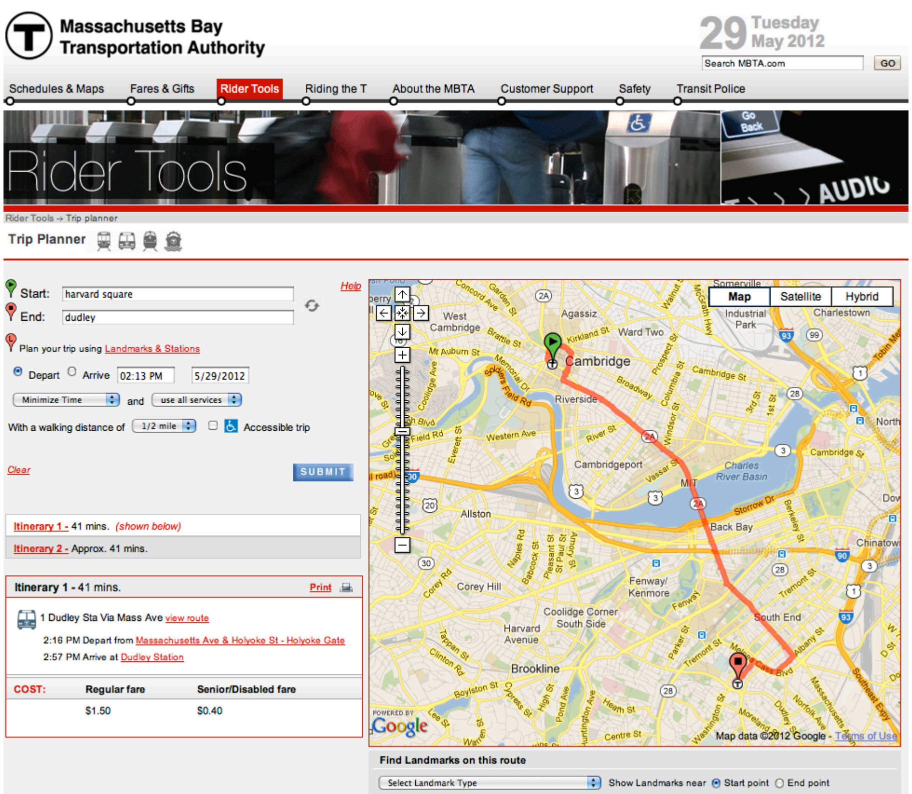


Figure 6 Online trip planner for Boston's MBTA

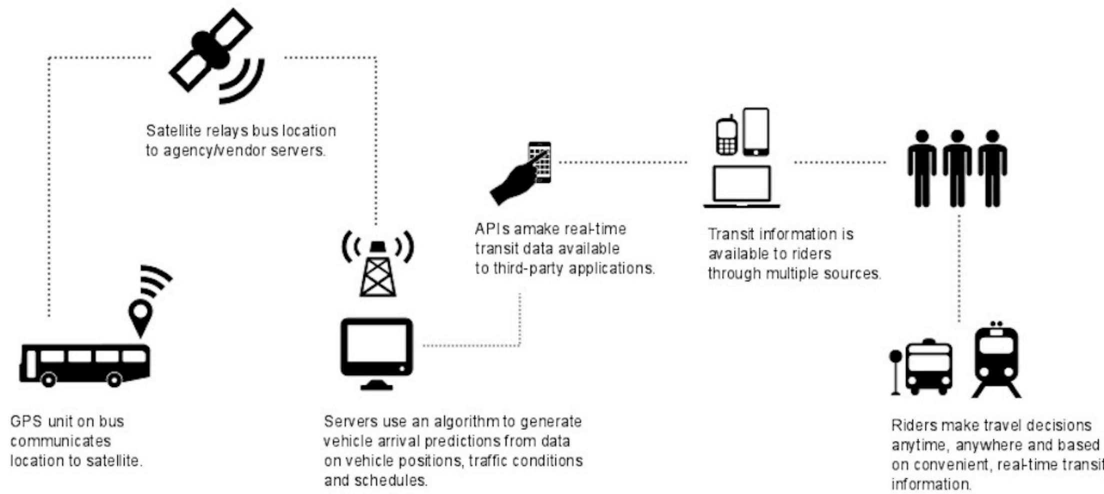


Figure 7 Technological systems used to generate real-time transit data. Diagram by author. All symbols from thenounproject.com collection.

Trip planning software evolved static schedules and routes further by allowing users to customize transit information for their individual travel needs. For New York's MTA, for instance, the agency's online Trip Planner service improved upon an existing telephone hotline by presenting riders with customized itinerary options for travel by bus, subway and walking. Yet, like printed and online schedules and routes, trip planners presented static information that did not bridge the gap between a system's intentions and actual on-the-ground conditions at a particular point in time.

The deployment of Automatic Vehicle Location (AVL) systems in the late 1990s solved this problem by enabling agencies to dynamically track the real-time location of transit vehicles through global positioning systems (GPS) for buses and signpost transmitters for fixed route systems like subways. Agencies used AVL for internal management purposes: to improve operations by monitoring and adjusting service performance and providing vehicle locations for emergency management, to name just two uses.

The type of digital data required to generate this information included geographic coordinates for transit stops and stations, geographic shapefiles of transit routes, and the geographic positions of vehicles. Vehicle position information and arrival predictions were complex pieces of data and constituted a composite of several values, including the precise location coordinates of a vehicle, the relation of a vehicle's location to the static schedule, and assessments of traffic conditions, among other data points (see Figure 7). For buses, where locating vehicles was particularly complex due to constantly shifting on-the-street conditions –due to traffic conditions, weather, passengers, and even drivers themselves – vendors such as NextBus and Clever Devices began to offer real time bus location and arrival prediction services to agencies, substantially lowering barriers to adoption of AVL technologies.

Agencies began to disseminate the real-time location and arrival of buses and trains to riders as a way to further improve their level of service to customers. The most prevalent method for communicating this information was through electronic signs located at bus stops and subway platforms. Once transit riders arrived at a stop or station, they would have a sense of how long their wait would be for the bus or subway. Starting in 2002, the TriMet agency in Portland made available real-time Transit Tracker information on its website for *pre-trip* planning purposes.[9] Knowing when to expect the arrival of a bus or train before heading to at a stop or station constituted a significant leap forward in disseminating travel information for transit systems.

2.2. Technologists demand transit data

Examining the process by which select transit agencies in the United States adopted open data strategies reveals that publicly disclosing raw operations data was not a foregone conclusion. Agencies did not anticipate that they would eventually share the raw data feeds designed for internal use with the public.

As discussed above, agencies were focused on installing their own infrastructure to deliver real-time train and bus arrival information to riders at stations and stops or on their agency websites. Electronic countdown clock signs that displayed the number of minutes until the next train were installed on station platforms as early as 2001 in Washington D.C.'s Metro system (i.e. WMATA). In Chicago, to comply with the American with Disabilities Act, the CTA outfitted its buses with GPS units so that electronic signs and audio could announce the approaching stop. The New York MTA's effort to install countdown clock signs in its subway stations is ongoing as of May 2012, with 177 stations (out of 468 stations in the entire system) delivering next train information to customers.[10] Most agency websites also featured system schedules, service alerts and trip-planning software (See Figures 4-6). Of the agencies studied for this report, only Portland's TriMet and Chicago's CTA offered real-time information through agency-built website applications prior to the diffusion of open data-based transparency.

These early dissemination initiatives relied on agencies to design, build and manage the interfaces through which riders accessed transit system information. But with the increasing adoption of mobile phones that could browse the Web, some transit riders began to visit agencies' websites for information while on-the-go. A small portion of these riders had enough technical skills and motivation to tinker with available online transit information in order to customize it for their particular needs – for instance, focusing on a few transit routes and optimizing displays for mobile phones. These computer programmers

[9] Transit Cooperative Research Board (2003) Pgs. 8-14. See TriMet's Transit Tracker at http://trimet.org/arrivals/index.htm

[10] A status update on the MTA's installation of countdown clocks is at http://www.mta.info/countdwn_clocks.htm

pulled desired information off transit websites, reworked the data into more user-friendly formats, and adapted it for different mobile platforms.

In Portland and Chicago, two tech-savvy individuals had connections inside the transit agencies and encouraged opening access to the raw data feeds through public APIs. Portland's Chris Smith was a local software developer and transit advocate with the group Portland Transport. In 2006 he started "screen scraping" TriMet's online Train Tracker service for real-time arrival information. Smith used the data to build the Transit Surfer mobile application to improve his own commute. Three different transit lines located a few blocks apart linked Smith's neighborhood to his job downtown and Transit Surfer allowed him to compare arrival times for the three bus lines on his smart phone.[11] This approach customized and improved upon TriMet's Transit Tracker web service, which at that time offered information for one transit stop per query and was optimized for websites, not mobile phones. Smith showed Transit Surfer to people he knew at TriMet, pitching for an open API to facilitate the use of transit data by other third-party applications. Eventually TriMet asked Smith to begin experimenting with a prototype API.

In Chicago, Harper Reed, a local software developer and entrepreneur, built his own API for the CTA's Bus Tracker real-time information website in October 2008. Reed wanted to help others build mobile applications to access real-time bus arrivals in a faster and more intuitive way than what was available on the agency's website. In an interview with Chicago's *A.V. Club* Reed recalled:

> The big thing was that my wife used [CTA] Bus Tracker, and it was kind of a pain in the butt. She had an iPhone, and the interface to it just wasn't necessarily friendly to the iPhone. It wasn't terrible, but it was enough where maybe she said one day, "This website blows." I just looked at it and was like, "How can this be better? There has to be a way to make this better."[12]

On a wiki, Reed documented how he reverse engineered the CTA's BusTracker website to build an API and wrote a blog post about it.[13] Almost immediately, his friends and other local developers began using Reed's API to build their own mobile apps for real time transit data in Chicago. Reed and his friend Dan O'Neil, also a civic-minded tech entrepreneur in the city, approached the CTA with their unofficial API, encouraging the agency to adopt a more formal, open approach to their data.

As software developers continued to build applications with unofficial versions of transit data, agencies began to see that by not making available official and

---

[11] See Chris Smith's list of favorite trips at http://tsrf.us/cps.html. Built for the earliest generation of smart phones in 2006, the information display is in HTML text only (the iPhone was released a year later, in 2007).
[12] Juliano, Michael (2011).
[13] Reed, Harper (2008) "Chicago Transit API." Also, see the "unofficial" Chicago transit API on GitHub at https://github.com/harperreed/transitapi.

timely datasets proactively, transit riders who used third-party applications would risk consulting outdated or erroneous information. Early adopters of open data like TriMet, the CTA and Boston's MBTA, also recognized that in expanding the number of outlets for their transit information, independent software developers were creating broad public benefit to riders at little cost to the agencies themselves.

In Washington and New York, organized groups of technologists and bloggers began to pressure the late-adopting agencies – Washington's WMATA and New York's MTA –to release schedules, routes and real time location feeds in 2009. These groups noted the precedent set by other transit agencies that had embraced open data and the accomplishments of independent developers in those cities.[14] Also, the presence of countdown clocks in these cities' subway stations raised developers' expectations that WMATA and the MTA already had the desired data and that releasing transit information to the public would be only a matter of pushing an 'export data' button. Though agency leaders, like the MTA's new Chairman and Chief Executive Jay Walder, understood the benefits to releasing agency data to the public, they faced institutional, technical and even legal challenges in doing so.

2.3. Challenges for agencies in opening up their data

When agencies first adopted intelligent transportation systems like AVL and GPS tracking for managing trains and buses, they had not anticipated the notion of independent, third-party agents working with operations data beyond the walls of their organizations. In meeting the public demand for disclosing data, agencies faced three main obstacles:

1. Institutional culture: at the time when technologists began to demand access to operations data, transit agencies as institutions were not particularly well-positioned to innovate in this space. Many agencies were experiencing significant financial constraints, which in some cases created pressures to reduce service and raise fares. (The Chicago, Washington and New York agencies initially explored ways to monetize their data before agreeing to release it to the public at no cost.) Furthermore, as public agencies, transit organizations were hesitant to release operations data that was not complete and of a very high quality, lest they jeopardize public trust. Public data releases could also expose performance difficulties. The tasks of pulling data out of existing software systems, getting it into a user-friendly format, and interacting with data users like developers also required collaboration between different departments within agencies that may not have had occasion to work together previously. And finally, as a new responsibility, public disclosure of operations data was not within the domain of any

[14] The Washington D.C. blog Greater Greater Washington sustained a years-long advocacy campaign for WMATA to publicly release data and incorporate schedules and routes on Google Transit. In New York, the non-profit group OpenPlans organized technologists to persuade the MTA to open its data through the NYtransitdata.org website.

previously defined role for agency employees. Those individuals who championed open data initiatives within their agencies did so out passion for the cause and on top of their other day-to-day responsibilities.

2. Data ownership: in certain cases, agencies did not own their own data. Many transit agencies contracted with private vendors to supply the technologies needed to plan routes, schedules and generate real-time location and arrival information. Under the terms of some contracts, agencies did not have the right to share the data generated by those systems with third-parties, meaning they were forbidden from disclosing operations information to the public. Agencies were also faced with patent-infringement lawsuits from ArrivalStar, considered by many to be a "patent troll," claiming rights to the ideas behind vehicle-tracking technologies. Agencies like the MBTA and MTA have settled with ArrivalStar, and must pay license fees either directly to the company or through their vendors for the use of vehicle-tracking technologies.[15]

3. Data format: even if the agencies did reserve the right to share their data with third-parties, schedules, routes and real-time data feeds did not exist *a priori* in a format amenable to third-party use. The GTFS data standard for schedules and routes facilitated translating this static data into a standardized, machine-readable format. But without a standard for real-time location feeds, agencies had to create datasets for public disclosure on top of existing operations management systems.

At the point when software developers began to request agencies' raw operations data, transit agencies had already been working with Google to integrate schedules and route information onto Google Maps as a complement to driving directions. Efforts to extract these data from existing technology systems, format it into the GTFS data standard, and share it with a third-party raised the above issues around data ownership, data quality and institutional constraints for the first time.

For transit agencies, integrating static timetable information with Google Maps via the GTFS standard was a critical precursor to the adaptive innovation of releasing this same data to the public. Forward-thinking transit agencies saw that Google was providing what amounted to online trip planning software for free and expanding the sources of transit information for customers. As David Barker, the Director of the MBTA's Operations Technology department commented: "our schedules are the best advertisement for the T that we have."[16] Agencies therefore had an incentive to pressure the few vendors in the industry to allow their proprietary data formats to be modified into the GTFS standard for use by Google. Once reformatted, GTFS was also a convenient format for third-party developers, who repurposed the data for use by riders in a variety of formats, providing even more diverse interfaces for schedule and route information.

---

[15] For further details see Snyder, Tanya (2012) and Badger, Emily (2012).
[16] Author interview with David Barker.

The innovative efforts of Portland's TriMet to expand the sources of transit data paved the way for these developments. Bibiana McHugh, TriMet's IT Manager of Geographic Information Systems, thought that transit routes and schedules should be as easily available online as driving directions. She approached Google, Mapquest and Yahoo with her idea, offering TriMet as a partner to add transit information onto existing online maps. McHugh received a response from Chris Harrelson at Google, who had already been working on integrating transit data into Google Maps as his "20 percent time" project.[17] Since TriMet had a centralized database for its operations information, McHugh and her husband Tim (the agency's Chief Technology Officer) were able to quickly pull the necessary temporal and spatial data elements, such as route numbers, route shape files, stop locations, and stop times. TriMet's timetable data became available through Google Maps on December 2005. TriMet and Google then worked together over the following nine months to develop a data standard for transit schedules, which became the General Transit Feed Specification (GTFS).[18] By the end of 2006, TriMet had made their GTFS file public under a Creative Commons license.[19]

TriMet was able to move quickly to expose its data to the public because it did not face the challenges faced by the other four transit agencies examined in this study. TriMet owned and managed its own information technology resources and, consequently, did not rely on outside vendors to generate operations data for its transit system. Institutionally, TriMet was well positioned to see the value of, and act upon, the notion of "creating more opportunities for easy access to transit information and for exposure to it."[20] Its information technology, communications and marketing functions were all housed within the agency's Communication Technology department, likely facilitating decision-making about such an innovative approach to delivering customer information. Further, TriMet employed in-house developers and therefore had the capacity to build its own services and tools, including a real-time arrival prediction algorithm, which precluded the need to contract with a vendor for real time feeds of transit information. Finally, the agency's lawyers were well versed in open source licensing options and were able to craft terms of use for the data that insulated the agency from risk. With an understanding of how information technologies and customer information were coming together, and without outside licensing constraints on its operations data, TriMet was able to innovate an open data strategy for the disclosure of transit information. Like the local transit advocate and civic innovator Chris Smith, TriMet's Bibiana McHugh had been lobbying from within the agency to expose Portland's transit data to the public as well.

[17] Ross, Matthew (2010).
[18] The GTFS data standard was originally called the Google Transit Feed Specification.
[19] A standardized, open copyright license that allows sharing of content. See http://creativecommons.org/licenses/
[20] Tim McHugh quoted by Antrim, Aaron (2008).

2.3. Diffusion of innovation: how and why did open data spread?

The open data strategy pioneered by Portland's TriMet spread to the largest transit agencies in the United States within the span of five years (see Figure 1) due to the convergence of five factors:

1. A *de facto* data standard offered by the GTFS format facilitated the process for agencies to integrate schedules and routes into Google Maps (now Google Transit), and for broader public disclosure of those same datasets;

2. Demand from technologically-savvy, networked transit riders for customized transit information due to the wide adoption of SMS-enabled cell phones and location-aware smartphones, which enabled riders to view real-time information while on-the-go;

3. Software developer communities who were eager to learn how to code mobile applications and sought available datasets to develop technical skills, and improve and expand access to transit information;

4. Agencies willing to adapt intelligent transportation systems used for internal management of operations into data formats suitable for public disclosure; and,

5. Open data champions who built networks within agencies to share experiences and seek advise on technical and policy aspects of data disclosure.

Integrating schedules and routes into Google Maps was an important motivation for agencies to publish GTFS files since it presented a free alternative to the expensive, online trip planners in which larger agencies had invested. As a private-sector player outside the transit industry, Google offered an alternative source for transit information through a free, timely, data-driven, and consumer-oriented product. Smaller agencies particularly benefited from this option. Even prior to Boston's MBTA information being available through Google Maps and to the public, Joshua Robin and Chris Dempsey at the Massachusetts Department of Transportation worked with the states' Regional Transit Agencies (RTA) to generate and publish GTFS data files for schedules and routes in areas like Worcester and Cape Cod. Larger agencies with existing trip planners were initially reluctant to do the same, however. In part they were hesitant to work with a data format developed by someone else, distrustful of Google's potential profit motives, and fearful that having their information on Google Maps would compete with their own websites' customer-information tools.

As riders began to access transit information through Google, they pushed the larger agencies to follow suit. "Customers wanted it, and saw the benefit of it, and would push General Managers on it," said Bibiana McHugh of TriMet.[21] In Portland and Chicago, civic-minded technologists who already had connections within transit agency administrative structures were able to lobby for an open data approach. For the late-adopter agencies, they eventually released datasets in response to pressure from local communities of software developers and bloggers. Eventually, transit managers at other agencies approached Bibiana and Tim McHugh with the challenge: "My General Manager wants us to be the next agency on Google Transit, how do I do it?"[22]

As the innovator in this space, TriMet's approach became the model to follow for other agencies. A key aspect of TriMet's disclosure model was that the GTFS files the agency developed initially for use by Google, were not handed over exclusively to Google. Bibiana McHugh made a point to establish that transit information was a public good, and as such, it should be accessible to anyone who was interested in working with that data. Also critical to the diffusion of open data among transit agencies was the strong advocacy and network-building work done by Bibiana and Tim McHugh.[23] They argued convincingly for an open approach to distributing transit data. In the third-party transit applications built by local Portland developers, they also had a tangible proof of concept that an open data approach benefitted agencies and riders alike.

Another benefit of having a common GTFS standard for transit data was that it sped up the rate of innovation and product creation by software developers. Community-produced tools emerged on how to use GTFS, including data libraries and message boards. A common data standard also enabled open source programmers to share and adapt each other's software code. This process delivered more choice to transit customers and extended the capacity of cash-strapped transit agencies to serve their customers' needs.

By the end of 2009, four out of the five agencies we examined had released static schedules and route information to the public: Portland, Boston, Chicago and Washington. Over the course of the next two years, all of these agencies made public schedule data and, to the extent available, real-time feeds of vehicle location or arrival predictions. Working closely with communities of developers, over time the most committed agencies to transit transparency expanded the scope of the data they released and improved the quality of the information they shared with the public.

[21] Author interview with Bibiana McHugh.
[22] Author interview with Bibiana McHugh.
[23] For example, see Tim McHugh's slides from the 2008 APTA Trans|Tech conference.

3. Communities of Transparency

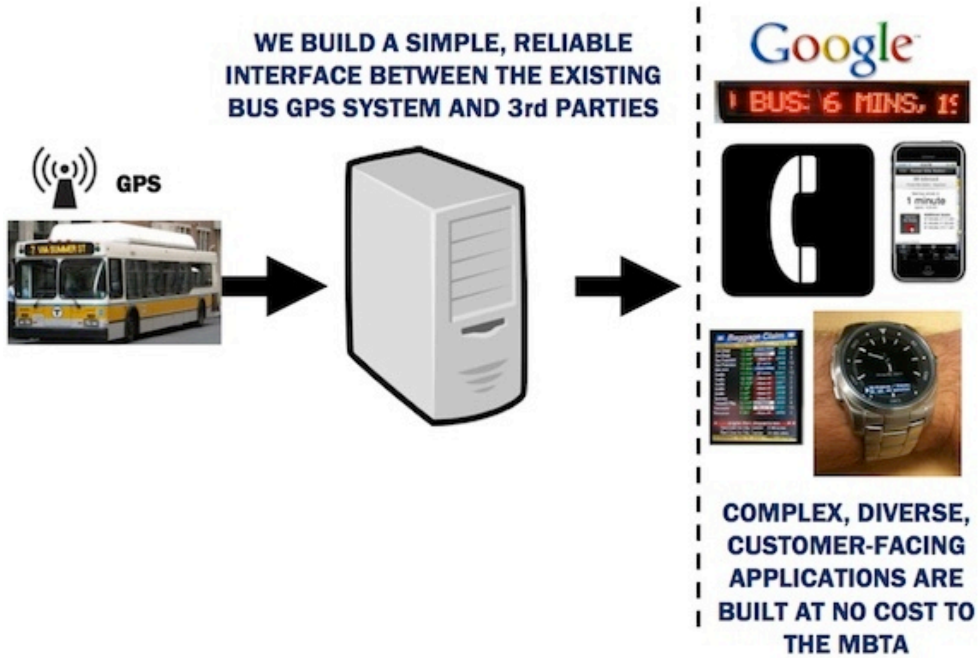


Figure 8 Transit data disclosure ecosystem. Diagram by J. Robin and C. Dempsey of MassDOT.

Driving the diffusion of open source transit disclosure (Figure 1) and critical to its success were the evolving communities of transparency that worked collaboratively to identify valuable datasets, refine data quality, and develop customer-facing digital applications for transit. This collaborative process involved three sets of actors that continuously interacted with each other: key staff at transit agencies who drove the discloseure of information, software developers as the information intermediaries who built digital tools for accessing transit information, and transit riders as contributors of ideas and feedback to improve the data and further refine the information tools. The result was a marketplace of third-party applications that offered many more outlets for transit information than could have been provided by transit agencies themselves.

Demand by technologists to access the same raw datasets and feeds used internally by transit agencies signaled a latent desire to customize the content and format of transit information beyond what agencies were already providing to riders. Once agencies publicly released GTFS files and built official APIs, independent software developers created applications optimized for multiple interfaces (mobile phones, desktop computers, tablets, electronic signage) and platforms (iPhone, Android, etc.), with diverse content and in various formats (Figure 6). They offered riders the ability to access transit information from multiple sources, in real-time, anywhere and at anytime. This was a radical change from when riders had to rely on printed timetables and static maps that, at best, represented the stated intentions of a complex system that could be disturbed by traffic, weather, personnel issues and even riders themselves.

3.1. Disclosers: Transit agencies

Transit agencies implemented transparency strategies for disseminating operations information more broadly by being responsive to demand for disaggregated data by local communities of technologists. Early adopter agencies in particular – Portland, Boston and Chicago – adapted their approach to customer information by complementing supply-side efforts (discussed in 2.1 Transit Information) with public disclosures of raw data. This shift from a closed, supply-side strategy to an open, demand-driven one occurred iteratively, in communication and collaboration with software developers who articulated their interests, motivations and needs in working with transit information.

Implementing the disclosure of disaggregated data required transit agencies to:

- <u>Convert transit data into formats suitable for public use</u>. Agency engineers had to tackle the task of pulling data from their internal management systems – which were not designed to generate raw data for third-party use – and converting these into formats for public use, such as the GTFS data standard for schedules and routes, and a real-time feed accessible through an agency's API. Agencies who had vendor contracts to produce real-time location and arrival information had to work with those vendors to open up those particular APIs, as was the case in Boston and Washington with NextBus and Chicago with Clever Devices.

- <u>Improve data quality to ensure accuracy and reliability</u>. The many elements that constitute transit data – geographic locations of stops, stations and routes, timetables, and arrival predictions – had to be accurate in order to generate reliable and trustworthy information that riders could use to make effective travel decisions. Agencies refined data quality through pilot deployments on a few bus lines (Boston's MBTA and New York's MTA), by releasing data in clusters once data quality issues had been resolved (in Chicago by bus sheds), and by generating new data altogether (Washington WMATA's need to geocode all bus stops in the system) prior to disclosure. Once data had been released, software developers played an important role in notifying agencies when data quality problems surfaced in the course of programming or use of their applications. Eventually, the most active and dedicated transit programmers also became beta testers for new data feeds.

- <u>Build interfaces for public data access.</u> Agencies established Developer Center pages on their websites to address third-party users of transit data. They provided two different ways for third-party developers to access transit data online. For static information – such as schedules and route locations – users could download GTFS files directly. Real-time location and arrivals information, because of its dynamic nature, required access to a continuous stream of feed messages, which could be accessed directly through APIs.

- Showcase third-party applications. Agencies also added a page to their websites to showcase the customer-facing applications built with transit data by independent developers. Careful to not endorse any particular applications, agencies included disclaimers such as the one below by the MBTA, which states:

> These apps are not made by MBTA, and MBTA does not sell or license the apps. They are written by third parties unless otherwise noted. MBTA shall not be held responsible for the content of third party websites or any issue arising from the use of third party applications. MBTA neither endorses any third party products listed here nor makes any guarantees or representations as to accuracy or reliability. Proceed with care and understand any usage changes that may apply to you. MBTA reserves the right to remove/add applications listings without notice.[24]

- Establish channels of communication with data-user communities. Agencies communicated with developers through online and offline methods. Online, Google's discussion list tool, Google Groups, presented a convenient forum for agency representatives and developers to exchange ideas, announce any data changes, and troubleshoot data quality problems. Face-to-face meetups, such as the series of developer meetings hosted by Boston's MBTA during the different stages of their data release effort (see Figure A1 in Appendix III) were important for building trust and a sense of shared purpose between the agency and its data-user community.

Because of the very different expectations, cultures, and orientation of transit agency personnel versus the software developer community, fostering relationships between the two groups was a critical ingredient for sustaining interest in, and improving, the process of transit data disclosure. The approach to developer relations by Boston's MBTA provides an illustrative case of how communities of transparency work collaboratively between information disclosers and information intermediaries.

The first meeting with local developers convened by MassDOT's Joshua Robin and Chris Dempsey in August 2009 involved a draft document that laid out "Relationship Principles" between the public agency and the prospective data-user community (Figure 7). The tone and language of this draft document revealed a misalignment of understanding between the agency and the

[24] MBTA Apps Showcase webpage, http://www.mbta.com/rider_tools/apps/

developer community, which was eager to work with transit data. For instance, the first line stated: "Don't abuse our data or our resources." (Figure 7) Laurel Ruma, then O'Reilly Media's open data evangelist, pointed out that, "If [developers] wanted to they could have crashed those servers a long time ago. But they haven't."[25] Ruma offered to act as a facilitator, encourager and mediator between MassDOT and the developers by redrafting the "Relationship Principles" in more conciliatory language that respected both agency and developer interests. The first line of the redrafted "MassDOT + Developers – Relationship Principles" document instead read: "MassDOT has limited resources and money. Respect the resources of MassDOT's partners and vendors who make data available to the developer community." (Figure 9).

These principles helped to manage developers' expectations about the MBTA's new approach to disseminating information to riders. It laid out the transformative potential of the agency's open data effort, but also its tenuous and uncertain status within the organization. Perhaps most importantly, this document established a common purpose for the public agency and third-party developers. This common purpose, and the mutual trust borne out of an understanding of shared risk around this endeavor, became the principal building blocks for future collaborative efforts around disclosing and working with transit data to produce customer-facing applications.

[25] Author interview with Laurel Ruma.

DRAFT EOT + Developers – Relationship Principles

+ Don't abuse our data or our resources
 - Doing so defeats our mission long term
 - We are looking to develop an API in the future – What might an API for EOT Developers look like?
 - Let us know if you have questions when possible before a problem occurs relating to our data. e.g. refresh rates on data.

+ We want people to be successful building applications
 - Don't forget point number one.
 - Benefiting from your work is fine with us.
 - Building applications does not increase the likelihood you will be awarded a contract by EOT now or in the future.

+ This is an experiment.
 - That means things can change, and probably will.
 - We will do the best we can to work with you – but EOT and state government are at times a volatile environment

+ Communicate
 - You with us: we will be as open and honest as we can be
 - Us with you: We are doing this in a new way and through a new type of relationship.
 - Tell us how we can do better, e.g. better use of Twitter or our Google Group?

Figure 9 Draft relationship principles by Robin & Dempsey for first developers meeting, Aug. 2009 (source: Laurel Ruma)

MassDOT + Developers – Relationship Principles

Please note: Any use of the Data on the Massachusetts Department of Transportation Developers Page acknowledges acceptance of MassDOT's Developer's License Agreement, as well as these principles:

- Respect Resources
    - MassDOT has limited resources (servers, admin/dev time, etc.) and money.
    - Respect the resources of MassDOT's partners and vendors who make data available to the developer community.
    - The community is participating voluntarily and also has limited resources.
    - Developers should not represent themselves as MassDOT or another state entity, and should respect trademarks (MBTA logo, state seal, etc.).
    - Please cite data sources.

- Succeed together
    - This project will showcase what is possible with limited resources.
    - Some projects might have broader appeal to other communities, so let's share.
    - Some projects could be profitable to the developer and that is fine.

- Expect change
    - Things can change. In fact, you should expect that they will.
    - MassDOT will make all efforts to keep data open once it's been released.
    - This is an experiment; success in building useful applications could help make it easier to open more MassDOT data in the future.

- Communication is key
    - Open, direct, and honest communication at all times.
    - Think about how the community can grow and evolve.

- Open, open, open
    - When possible, use open standards, formats, and software.
    - Be willing to share your successes.
    - Interact with community members in a productive manner.

- Legal
    - Building applications does not increase the likelihood you will be awarded a contract by MassDOT now or in the future.

Figure 10 Final version of developer relationship principles by MassDOT, Nov. 2009 (source: http://www.eot.state.ma.us/developers/downloads/RelationshipPrinciples_11-12-2009.pdf)

3.3. Intermediaries: Software Developers

Local programmers were the information intermediaries between agencies and riders, taking raw timetables, geo-codes and location feeds to generate a variety of information tools that transit agencies could not have imagined on their own. Independent, third-party software developers had various motivations for working with raw data to improve access to transit information. Below are the four most common motivations discussed by developers during interviews for this study:

- Improve their commute. "I wrote it for myself, I didn't expect it to be on the market," commented Chris Cieslak who built Chicago's Buster app.[26] Cieslak, like many other developers, worked with raw transit information in order to improve upon and customize agency-built information systems to facilitate theirs and their friends' commutes. Developers worked with transit data to solve their own problems and optimize their own access to information for specific purposes. These applications were custom-built software for a small community of users that overlapped closely with developers' own mobility needs.[27]

- Learn to program mobile applications. Transit data releases coincided with the broadening adoption of smartphones like the iPhone (which was released in 2007). Developers wanted to learn how to write applications for this new generation of powerful mobile devices, which were location-aware and featured touchscreens. Transit data was increasingly available for third-party use, and contained the temporal and geographic elements that coincided well with the capacities of smartphones. As a result, developers worked with transit data to learn how to code mobile apps.

- Civic innovation. Some developers were driven to build applications with transit information for civic purposes, believing that it would help people use transit more often and discourage the use of cars. Beyond these environmental sustainability objectives, some developers also believed that at least some transit applications should be free for users to access and download, given that transit information was a public good, which should be publicly available at no cost.

- Profit. Interviews with developers revealed that the *least* common motivation for working with transit information was to profit from building customer-facing applications. A few developers indicated that they set out to make money off their apps on Apple's iTunes App Store or the Android Market, recognizing the market value of more convenient access to transit information. Yet other developers who also chose to sell

[26] Author interview with Chris Cieslak.
[27] Clay Shirky (2004) calls these types of applications "situated software." They are "form-fit tools for very particular needs… [that] function well, because they are so well situated in the community that uses them."

their applications saw the extra money as a bonus. Another set of developers stated that they chose to give their applications away for free because it lowered user expectations around how applications functioned and avoided the hassle of managing money. Plus, software developers noted that they were already paid well in their jobs as computer scientists. Portland's Andrew Wallace explained: "If you charge even a small amount, people start expecting a whole lot. I already have a job, I don't want a second job."[28]

Harper Reed's experience in Chicago illustrates the above motivations well:

> I spent a lot of time reverse engineering the Google Maps mashup the CTA had, and I spent a lot of time trying to document what they actually did... So I showed it to some friends, and they were like, "Wow, that's pretty awesome," and we brainstormed what apps we could make. And then I was like, "You know, I don't want to make any apps. I just want to make it so it's easier for my wife to catch the bus." I put up a Wiki page that was just like, "Here's how you can do it yourself, and here's the API to get the bus tracker," and then a bunch of friends released apps.[29]

Reed was curious to learn how the CTA's BusTracker system worked, he sought to improve his wife's commute, and on top of that, established an API to help his friends and others build their own solutions using access to real-time bus locations.

3.4. End-users: Transit Riders

Transit riders also played an important role in the communities of transparency that coalesced around transit information. As users of third-party applications, they contributed to further improving these interfaces by troubleshooting software code, flagging data quality problems, and adding customized features. Mobile applications on the iTunes and Android apps stores are required to integrate a feedback feature for users. This feedback mechanism allowed transit riders to report bugs and problems with the applications back to developers. Transit riders could also use these feedback mechanisms to report problems with the underlying data that populated the apps. Developers filtered data quality complaints back to the agencies for resolution. In interviews, developers commented that end-users of transit applications did not hold agencies accountable for data quality problems directly. Apps users instead believed that when arrival predictions were inaccurate, it was the application itself that was not working well rather than ascribing problems to the raw data feeds. In a sense, third-party applications sometimes insulated agencies from data quality complaints.

[28] Author interview with Andrew Wallace.
[29] Harper Reed quoted by Juliano, Michael (2011).

Feedback mechanisms for third-party applications also became conduits for user-driven ideas, adding features to further improve the functionality of applications. Users in Portland suggested particularly creative features for mobile applications, including a snooze function to wake sleeping bus riders when they approached their stop or a flashlight function so bus drivers could see that people were waiting at stops during dark winter days. Further, people with disabilities asked developers to make their apps more accessible. For instance, the PDXBus iPhone application used audio to read out information on the phone's screen for vision-impaired users and displayed the bus number on the screen so that riders could hold up their phone for drivers to read once a bus approached a stop. Bus operators could then confirm whether a rider was about to board the correct bus. Because Portland has such a rich ecosystem of open data, developers also integrated data on bicycle parking locations, weather, and TriMet's new open source trip planner to their applications.

3.5. Marketplace of Apps

The result of the collaborative process between transit agencies as disclosers, developers as information intermediaries, and transit riders as end-users, was the growth of a vibrant marketplace of applications built around transit information. In marshaling the energy and creativity of local developers, transit agencies gained broader exposure for their systems' customer-facing information.

The strength of agencies' relationships with software developers proved a critical distinction between early-adopter agencies and late-adopter agencies, influencing outcomes in the number and variety of applications produced for each system (Tables 2 and 3). As discussed earlier, Portland and Boston have actively fostered strong relationships with their local developer communities, building trust and a sense of common purpose. Washington's WMATA, absent a sustained effort to be responsive to developer needs, has relatively anemic results in terms of the number and variety of third-party applications using their data.

Table 2 details the number of third-party applications built for each of the five transit systems examined in this report, ordered by the date of data release. The early adopting agencies tend to run smaller transit systems (as measured by average weekly ridership figures) yet have many more transit applications per rider than late adopting agencies. In Portland, there's one independent application for every 7,000 weekly riders whereas New York, which has the highest absolute number of applications, has such a vast transit system that the ratio of applications per rider is roughly one per 128,000.

In terms of overall usage of third-party applications by transit riders, the MBTA's interim General Manager Jon Davis noted that as of May 2012, "more than a hundred thousand smart phone users have downloaded apps that provide arrival time information for more than 180 MBTA bus routes."[30] As for the volume of use that a single application garners, at the upper end of the spectrum is an application like PDXBus in Portland. At approximately 30,000 downloads, PDXBus provides transit information to nearly ten percent of TriMet's average weekly ridership.[31] In Chicago, some 60,000 iPhone users had downloaded Chris Cieslak's Buster application for the CTA as of April 2011. Even applications built entirely for the use of individual developers have found public users. The myTriMet.com application built for mobile web browsing by John McBride gets a steady 20 visits per day during weekdays.

Table 3 summarizes the number of applications serving each transit agency by type of platform. Overall, the most popular platform is iOS, for use on iPhones, iPads and iPods. Applications built for Android phones and for mobile and desktop web applications are also population. Transit riders in Portland and Boston, which have the strongest developer relationship mechanisms of all the agencies, along with New York City, the largest transit agency by many orders of magnitude, enjoy the greatest variety of platforms on which they can access transit information. A critical distinction between the early adopter agencies in Portland and Boston and New York is that riders in the two smaller agencies can access real-time information for those entire transit systems (with the exception of the Green Line trolley in Boston). In New York, transit riders can access real-time data for buses in Staten Island, two buses in Manhattan, and one bus in Brooklyn. As a result, many applications built for New York's MTA offer static schedules and maps. Yet those developers that have already incorporated the MTA's real-time bus routes into their applications are well positioned to incorporate future data feeds due to the agency's use of standard data formats. The MTA expects to deploy BusTime on all routes in New York by the end of 2013.

[30] Quoted by Annear, Steve (2012).
[31] Author interview with Andrew Wallace, developer of PDX Bus, on September 30, 2011.

TABLE 2 NUMBER OF TRANSIT APPLICATIONS BY CITY AND AVERAGE WEEKDAY RIDERSHIP, ARRANGED BY DATE OF DATA RELEASE (sources: Average weekday ridership figures for TriMet, MBTA and CTA from APTA Transit Ridership report for first quarter 2012, representing bus, heavy rail, light rail and trolley ridership. Average weekday ridership figures for WMATA from *Metro Vital Signs Report* for March 2012 and for New York MTA from *MTA Finance Committee Meeting* memo for April 2012. Other information from agency websites' App Centers and author's research)

Agency	Portland TRIMET	Boston MBTA	Chicago CTA	WDC WMATA	New York MTA
Av. Weekly ridership (2012)	322,500	1,174,900	1,711,900	1,214,024	8,578,575
Number of apps (2012)	45	44	21	10	67
Ratio (apps/ridership)	1/7,000	1/27,000	1/82,000	1/121,400	1/128,000
Agency-developer relationship	Strong	Strong	Medium	Weak	Medium
Marketing push	Light	Light	Medium	Light	Heavy
Year of initial data release	2006	2009	2009/2010	2010	2011/2012

TABLE 3 NUMBER OF THIRD-PARTY APPLICATIONS BY PLATFORM, 2012
(Sources: agency websites' Apps Centers, current as of June 2012. Note: Certain applications are counted more than once because they are available on more than one platform, i.e. iOS and Android and on a web browser)

Agency		Portland TRIMET	Boston MBTA	Chicago CTA	WDC WMATA	New York MTA
Mobile	iOS (iPhone/iPad)	14	26	10	4	53
	Android	10	10	7	4	23
	SMS	3	3	-	-	4
Web apps		12	12	5	1	7
Digital signs		3	1	1	-	1
Other	(Blackberry, etc.)	8	5	2	2	14

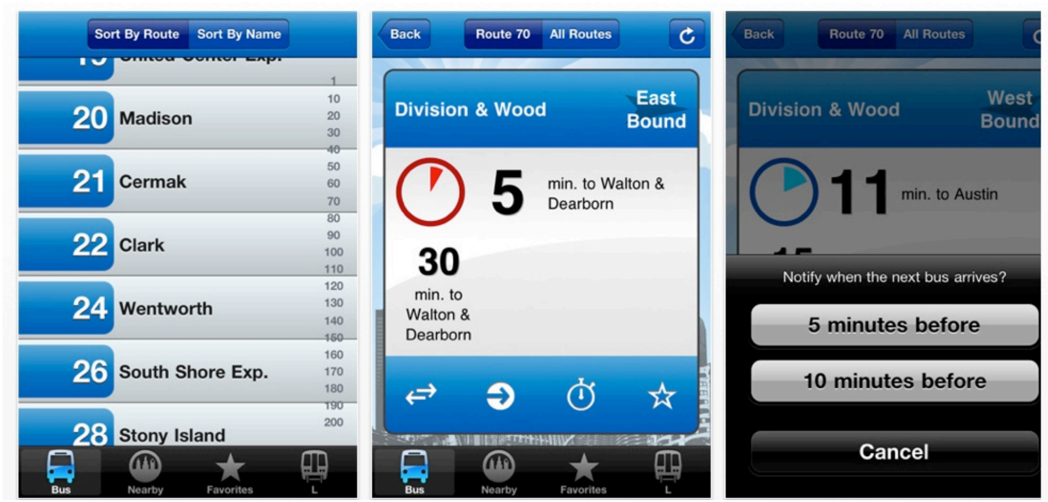


Figure 11 Buster application for Chicago's CTA transit system

Common data standards for schedules and routes (GTFS) and for real-time information (SIRI and GTFS-realtime) make it possible for a single application to be used for travel on any transit system that releases data in that standard. Adding additional transit systems to an application simply requires changing the base maps within the source code of an application. Several applications do this. For example, HopStop provides transit information for 78 transportation systems in the U.S. and abroad. Catch the Bus, built originally for Boston's MBTA, eventually expanded to include San Francisco's MUNI system and the MTA's B63 route in Brooklyn, among others. But an application that serves more than one transit agency is unusual in the ecosystem of third-party applications that has evolved around transit data.

As "situated software," Clay Shirky's term for digital tools that are built to satisfy very particular needs tailored to a community of users, transit applications tend to be customized for each system it serves. Chris Cieslak, the developer of Chicago's Buster application, used the CTA's design language in laying out the elements of his application (Figure 10). Cieslak noted the importance of local knowledge in understanding the possible "use cases" for an application: "If you haven't been there, you'll have no idea how people will use the app."[32] Other developers and transit managers echoed this same sentiment. When announcing a new real-time data feed for the MBTA's commuter rail, Josh Robin, Director of Innovation, told developers: "We feel strongly that local developers are where it's at. You know our riders, you know our system."[33] George Schneeloch, who developed Boston Bus Maps for Android phones, said that he hesitated to work on applications for other cities because he lacked local knowledge and wanted to limit the amount of time he worked on what was essentially a side project for him. Nevertheless, Schneeloch noted that since the code for his application was open source, other developers could use his Boston

[32] Author interview with Chris Cieslak.
[33] From author's notes.

Bus Maps code to build applications for other cities, adapting and improving the code to serve other localities and use cases.[34]

[34] Author interview with George Schneeloch.

4. Outcomes

At this time, many of the outcomes of transit transparency efforts are difficult to assess. This section discusses the outcomes we have been able to capture as a result of transit agencies' open data strategies. First, we recognize that the process of opening up operations data marks a new model for customer information delivery on the part of agencies. Certainly an open, demand-sensitive approach to providing transit information disclosure on the part of agencies represents a significant shift in strategy for disseminating customer-information and an important outcome in itself. Second, while there is little available evidence that data disclosure is driving improved performance by transit agencies,[35] we present a few examples of how individuals have used transit data for analyses and visualizations of transit system performance. Finally, to address questions regarding the influence of real-time data availability on rider effects, we discuss some preliminary results from onboard customer surveys in Boston and New York City. Our initial findings support prior research that mobile access to real-time bus arrival information decreases the perceived and actual wait times for riders. Further, improved access to transit information gives riders greater discretion over their time. However, it is too early to tell whether reduced wait times translate into increased satisfaction with overall transit service and whether information is enough to produce those effects independent of operational improvements like increased service frequency.

4.1. A New Model for Customer Information Delivery

Starting in late 2005, forces from within transit agencies and outside of these organizations coalesced to produce an entirely new approach to disseminating information about transit systems to customers. In contrast to traditional, vertically integrated customer information strategies, efforts by transit agencies for greater transparency involved co-production of customer information delivery with local programmers. As discussed in the prior section, this process yielded a rich and dynamic marketplace of independent applications. As the MBTA's Joshua Robin explained, data disclosing agencies became "wholesalers of information."[36] Local software developers in turn became the "retailers" by designing the information delivery systems for transit riders.

This shift in strategy evolved through an iterative process, requiring a sense of common purpose between agencies and developers. In essence, agencies incorporated developers into their data supply and quality-assurance process. Developers communicated their data needs to agencies. Agencies worked to generate and shape data into formats suitable for disclosure. Developers would then beta-test datasets to improve quality. And as agencies made more data available, developers updated and refined their applications to reflect these improvements. Over time, transit agencies expanded their disclosures from static

[35] In Chicago, however, real-time bus data was released only after the agency improved on-time arrival metrics, according to Professor Vonu Thakuriah at University of Illinois at Chicago. For more, see the Chicago CTA Case Study.
[36] Author interview with Joshua Robin.

schedule and route information to raw feeds of dynamic, real-time arrival predictions. The dissemination to customers of this information went from relying on a single agency website and expensive digital hardware installed by agencies, to the widespread adoption of portable digital interfaces by riders themselves, who could then access mobile applications for transit after leaving home and before deciding to wait at a bus stop or enter a train station.

The process of data disclosure highlighted for some agencies constraints embedded within the vendor contracts on which they relied to manage their intelligent transportation systems. Portland and New York, the smallest and largest of the agencies examined in this study, adopted open source tools to manage their transit systems and generate the necessary public-facing data, particularly for transparency of real-time location feeds. These non-proprietary tools were built in-house or adapted from existing source code, allowing these agencies to ensure rights to the data and to build the internal technical capacity needed to manage systems from within the organization.

New York's MTA benefited from its late-adopter status by being able to learn from other agencies' prior experiences and becoming a leader in implementing open source software and hardware systems for real-time data disclosure. It leapfrogged other agencies in innovating an open source approach to installing location tracking hardware on buses. Instead of working with "black box" solutions where a vendor installs, manages and maintains the GPS units and software for bus fleets, the agency implemented an innovative modular system using off-the-shelf components: GPS for location tracking, a computer for collecting data, software for conveying information to the public, and wireless communications for routing data. The MTA required the different vendors from which it procured the components to have a standard interface so the modular system could operate as a whole. Investments in these components were happening anyways as the agency upgraded its payment systems. By repurposing the payment system to also collect and disclose bus locations, the agency "essentially got real-time information for free."[37] The MTA's open source approach was an impressive accomplishment for an agency whose former Chairman and Chief Executive Jay Walder admitted in early 2011 that, "The MTA doesn't do technology. It has a Star Wars strategy instead, where technology only happens in a galaxy far, far away."[38]

4.2. System performance and accountability

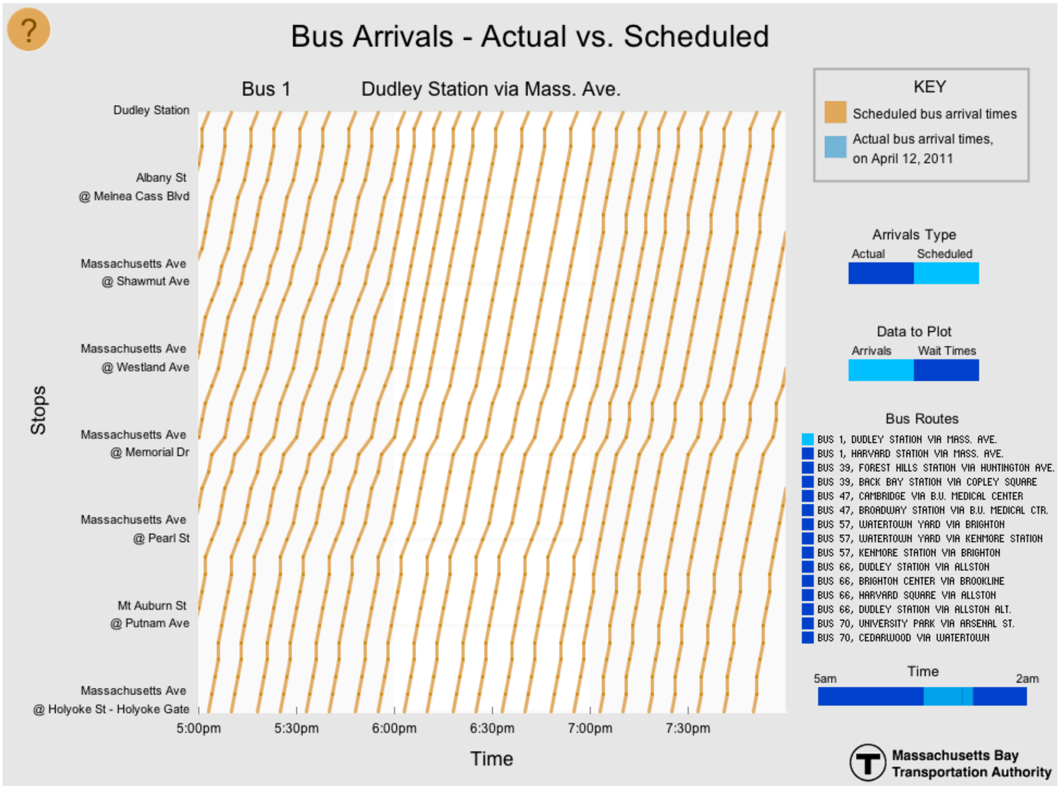
The transparency action cycle framework (Figure 2) used in this study conceptualizes how information becomes embedded not only in the actions of information users but also in those of the disclosing entities. One of the initial hypotheses for this study anticipated that the disclosure of disaggregated operations information would result in agencies improving the performance of

[37] The MTA's Michael Frumin at an MIT presentation, February 25, 2011. For further details, refer to the New York MTA case study.
[38] Jay Walder presentation at Harvard Kennedy School titled "Public Transit in New York," March 2, 2011.

transit systems as public users of the data generated their own performance metrics for accountability purposes. While we have not found systemic studies that examine whether disclosure of transit information is driving improved performance, Chicago's CTA worked to improve on-time-performance for buses in tandem with data quality. Anticipating public access to real-time bus positions, the CTA strategically improved performance in clusters, bus shed by bus shed, and staggered its release of real-time bus positions accordingly. Aside from Chicago's anticipatory performance improvement efforts, we know of only a few instances of individuals using transit data to assess the performance of transit systems. These have not been sustained or systematic analyses of performance, however.

In Washington D.C., former-*Washington Post* reporter Luke Rosiak provided a "window into performance" by comparing real-time bus locations to schedules on December 9, 2010. He found that 30 percent of buses were six or more minutes off schedule, while 18 percent ran within a minute of their schedules.[39] Two Harvard computer science undergraduates, Adrian Sanborn and Chioma Madubata, performed a similar analysis with MBTA data for April 12, 2011 (Figure 11).[40] They used interactive data visualizations to highlight the disparities between schedules and actual bus locations over the course of the day for several bus lines in Boston. The top image in Figure 10 displays scheduled runs along a bus route between 5 and 8pm. Each stop is displayed on the y-axis and time of day is on the x-axis; the lines connecting each stop and time of day represent individual bus runs. The yellow lines indicate scheduled bus runs and in the figure below they are regularly spaced, indicating that wait times between buses are scheduled to average approximately four minutes. The bottom image displays the actual trajectories of city buses (blue lines), logging the arrival time at each stop along a bus route. The visualization of actual bus arrival times clearly shows a pattern of bus bunching, when buses clump and arrive together, and uneven headways, where average wait times are approximately six to seven minutes.

[39] Rosiak, Luke (2010).
[40] Sanborn, Adrian and Madubata, Chioma (2011).



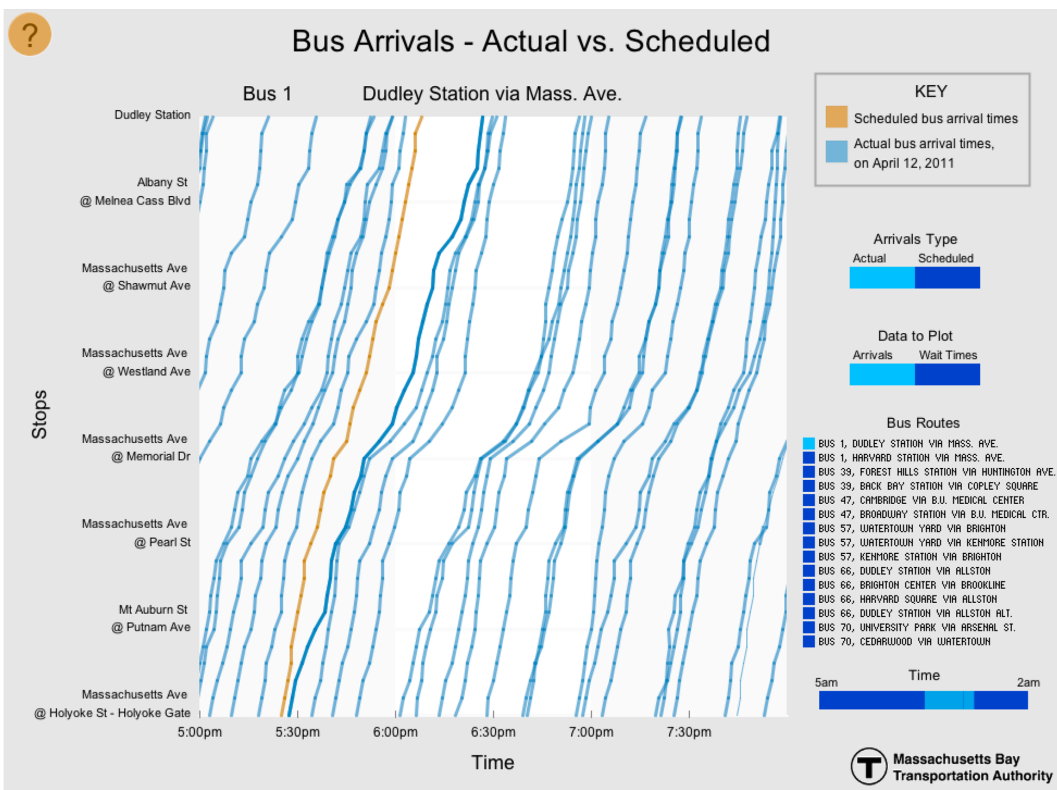


Figure 12 Comparison of scheduled bus arrival times (yellow lines) and actual bus arrival times (blue lines) for Boston's MBTA #1 bus using GTFS and real-time data for April 12, 2011. (by Sanborn and Madupta http://www.cs171.org/2011/projects/web/Sanborn-Madupata/applet/)

A few independent developers have also built amusing, yet useful, websites as performance dashboards. Harper Reed and Dan O'Neil in Chicago, who played an important role in convincing the CTA to establish a public API for its data, also built the "CTA Alerts "website to show how *well* the transit system works in their city. On a day-to-day basis, CTA Alerts monitors service alerts for the entire CTA system and provides a user-friendly interface for viewing alert details by bus or train line. CTA Alerts declares in a cheeky tone: "…this urine-soaked system of transit is darn good, as far as we can tell, and now we've got the data to prove it."[41] In Boston and Washington, the "How's the T?" and "How's the Metro" websites provide an inverse take on this same performance dashboard idea. These websites display the average wait time and longest wait time for each different train line in the system with assessments of each train line's performance phrased as: "looks just fine," "might be a little spotty," "doesn't look so hot."[42]

Perhaps we find so few efforts at assessing system performance by comparing transit schedules to actual arrival information is that once real-time data is available, schedules become a less important source of information. If at some point in the future, all riders can consult real-time location and arrival information along an entire transit network, the notion of "on-time-performance" may no longer be based on schedules, but rather on some expectation of service regularity that meets customer demand. It may be that with the availability of real-time information, transit riders are adapting their use of transit to a system "as is" rather than to the expectation set by schedules. Along those lines, the following section presents some initial findings of customer surveys that seek to capture rider effects of improved transit information delivery.

4.3. Rider effects

To examine changes in rider behavior and attitudes about transit service as a response to better access to information, we employed pre- and post-wave surveys of transit riders in two of our case study transit systems: New York's MTA and Boston's MBTA. These onboard surveys asked riders about their transit behavior and attitudes regarding service both before and after agencies made real time transit data available to the public. In New York the aim was to understand the effects of real time information on riders who rely on bus service along the B63 line in Brooklyn. In Boston, the focus was on the effects of real time information for riders of the Worcester and Newburyport/Rockport commuter rail lines. This section discusses some preliminary, high-level findings that inform future hypotheses to be tested upon completion of post-wave surveys in Boston and New York.

Preliminary survey results indicate that most transit riders have the technological tools to access digital information. We found very high rates of mobile technology use among transit riders. On the B63 bus line in Brooklyn, whose ridership is representative of the MTA's overall demographic for bus service, 88

[41] See http://www.ctaalerts.com/
[42] See Joey Brunelle's web applications http://howsmetro.com/ and http://howsthet.com/

percent of riders surveyed used cell phones, and 49 percent used smartphones. On Boston's commuter rail, 95 to 97 percent of riders used cell phones and 60 to 68 percent used smartphones.

When asked about use of real-time arrival information, results from the bus and commuter rail surveys are consistent. In New York, the MTA's post-wave survey on the B63 bus line indicated that over half of respondents had used BusTime's real-time bus location information, and that nearly half of those aware riders consulted BusTime on all of their bus trips. The MTA conducted this survey six months after introducing BusTime for the B63 route in New York, at the time the only bus line in the city with real-time information. These relatively high rates of awareness and adoption on the B63 so soon after implementation show promise for broader BusTime adoption in New York.

Surveys on Boston's commuter rail prior to the availability of real-time data on that mode of travel revealed that over 50 percent of riders were aware of the availability of real-time information for MBTA bus and subways. Further, 30 percent of respondents reported using the real-time information when traveling the bus and subway in Boston. These are notable rates of awareness and adoption for commuter rail riders given that bus and subway were not their primary modes of travel and that the MBTA did not deploy a marketing campaign around real-time data availability. Post-wave surveys conducted in June 2012 will reveal use rates of real-time information for commuter rail riders, one year after the information became publicly available.

When asked about riders' levels of satisfaction with service quality, bus riders on the B63 bus were least satisfied with wait times. In terms of the effects of real-time information on wait times, the MTA's post-wave survey found no statistical difference between wait times on the day of the survey for B63 riders using BusTime and riders on a control route without BusTime. However, when asked to report their *usual* wait times, BusTime users consistently estimated lower usual wait times than non-users of BusTime on the B63 line.

Another possible benefit of real-time information provision to bus riders is that information regarding the status of buses and trains provides riders greater discretion over how to use their time. Over half of the B63 riders who consult the MTA's BusTime bus arrival information reported spending some of their time waiting for the bus "elsewhere", that is, at a place other than the bus stop on the day they were surveyed. By comparison, among B63 riders who do not consult BusTime information, approximately one-quarter reported spending time elsewhere while they waited for the bus. This is a statistically significant difference in behavior between BusTime users and non-users.

These early findings support prior research that mobile access to real-time bus arrival information decreases the perceived wait times for riders. Watson et al (2011) found in Seattle that the use of mobile, real-time information by bus riders decreased their perceived and actual wait times. However, it is not yet clear whether better information improves riders' overall satisfaction with transit service or increases ridership numbers. A case study of real-time information ridership effects in Chicago by Tang and Thakuriah (2011) indicates a small increase in ridership, suggesting that people who already use the transit system ride more often due to better information delivering greater convenience. Our further research aims to examine these longer-term ridership effects of transit information disclosure.

Prior research findings and our preliminary survey results in Boston and New York allow us to formulate the following hypotheses about rider effects of transparent transit data:

- High rates of cell phones and smartphone use enable high rates of digital access to mobile transit data.
- The use of real time data decreases wait time as riders make informed, timely decisions.
- As riders are able to access and customize real-time information to make timely decisions about their travel, customer satisfaction with transit service may increase.

Forthcoming analyses of pre- and post-wave survey results on Boston's commuter rail lines and New York's B63 and Staten Island buses will provide insights into the above hypotheses.[43]

[43] These analyses will be carried out by Candace Brakewood, PhD student at Georgia Tech, in conjunction with researchers at the Transparency Policy Project.

5. Conclusion & Recommendations

The five case histories in this study show that transit agencies adopted transparency strategies for their operations data in order to improve customer service. Agencies created more opportunities for riders to access transit information through a process of co-production with local software developers who acted as third-party information intermediaries. Local computer programmers reshaped, reproduced and recombined the disaggregated, machine-readable electronic files made available by transit agencies to meet their own travel needs and those of diverse transit riders. The supply and demand processes of disclosure and uptake of transit data generated a rich and diverse marketplace of customer-facing digital tools and applications for riders.

This study finds that the public disclosure of transit information by agencies is a very successful case of open data adoption in the United States. The success of transit transparency offers insights into the elements that enable effective disclosure and delivery of digital information to the public in cases where there is a strong demand for that information, and where the disclosed information is available at the right place and time for users to act upon.

5.1. Why did transparency work so well for transit?

The case of transit data disclosure appears to exhibit several elements that prior studies have found to produce effective transparency.[44]

- An information gap existed that could be bridged by better public data. Making transit information available in disaggregated, machine-readable formats allowed third-party developers to work with the data to produce customized applications for their specific travel needs. This effect was particularly strong in the drive to optimize transit agency data for mobile access on smartphones and cell phones. In this sense, when transit agencies responded to developers' demand for data, they helped bridge two types of information gaps. One was the need to access transit information in mobile formats to enable travel decisions anywhere and at anytime. And the other was the need for real-time location information about transit vehicles, since "schedule information can be misleading if the system is not operating on schedule,"[45] as Figure 11 illustrated is often the case.

- There was a substantial demand for new data. Communities of technologists were the first to identify the need to access raw transit data in digital formats. In their responses, agencies targeted and engaged the right intermediary users for the disclosed information: computer programmers with the motivation and technical skills to work with raw data. The agencies with the most effective transparency outcomes were

[44] Adapted from Fung, Graham and Wei (2007) Pg. 174-175.
[45] Transit Cooperative Research Program (2003), Pg. 31.

the ones who developed strong mechanisms for developer relations. Agencies and developers worked together in an iterative way to refine data according to user needs and agencies' technical capacities. In effect, agencies integrated information-users into the supply of information and developed a process for the co-production of transit data with these "expert" users. Also critical to the ecosystem of transit data was demand for greater diversity and customization of transit information by end-users, riders without the expert skills to code with raw data feeds who contributed by providing feedback to troubleshoot data and ideas that added rider-friendly features to third-party applications.

- Communication of the information was practical. The demand by technologists for better transit data came about at the same time as the availability of sophisticated smartphones that enabled access to information at any time and anywhere. Transit information contained geographic and temporal elements that were well suited for the new capabilities of location-aware smartphones. Independent developers leveraged these advances in information and communications technologies to translate complex datasets into user-friendly applications, displaying transit information through interactive and dynamic maps, distances, timetables, and countdown clocks. The rich variety of high-quality applications available for use by transit riders responded was situated in an array of use cases that transit agencies on their own would not have had the resources to address in-house.

- There were consensus metrics for the information. The development of the GTFS open data standard for timetable and route information was an important baseline for other agencies, presenting a roadmap to information disclosure. The GTFS standard served as a consensus metric for information users. The existence of a common data standard facilitated developers' collective learning process for integrating transit data into customer-facing applications. Further, the diffusion of data disclosure across transit agencies and the broad takeup of GTFS data by third-parties served as a proof of concept for developing a standard around real-time data feeds. In its turn as innovator around open source approaches to generating and disseminating operations data, New York's MTA was the first to use the SIRI real-time data standard for its BusTime deployment in early 2011.[46] Later that year, Google introduced the GTFS-realtime standard, integrating Portland and Boston data feeds into Google Maps. As of mid-2012, New York's MTA is also beta-testing GTFS-realtime for its future release of subway data. It seems likely that real time data standards will also spread to other agencies, facilitating further use by developers and interoperability between transit agencies.

- Once available, transit data provided riders with useful information. The variety of formats and platforms offered by developers to riders made

[46] For a clear and in-depth discussion of real-time data standards, see Raschke, Kurt (2011).

transit schedules, routes, service alerts, and arrival information convenient and accessible. Transit riders were able to make pre-travel decisions about their transit use in ways that best fit their needs and capacities, at the right time and in the right places. Real-time arrival predictions in particular took the uncertainty out of transit, allowing riders to use their time with more discretion and reduce wait times at stations.

In essence, agencies got the data supply strategy right. Transit agencies' disclosure of operations data improved upon prior customer-information systems because: a subset of transit riders with programming skills were able to improve upon existing customer-information systems by customizing raw data to meet rider needs; the development of a data standard for schedule and geospatial information allowed quick adoption by transit agencies and data uptake by independent software developers; and the proactive engagement of local software developers by transit agencies fostered a sustainable community of use around transit data. Once applications were developed, the transit user community was ready to integrate the new information into their daily decision routines regarding travel and to provide feedback to developers on how to further improve the accuracy, format and scope of that information. It also helped that the type of information provided was highly salient: a few pieces of data told riders what they needed to know: where is the bus? Or, where is the train?

5.2. Lessons for other disclosure systems

From the disclosure experiences of the transit agencies examined for this study, we draw the four lessons to consider in designing future transparency systems.

Lesson 1: Identify the problem to be solved with better data. The critical questions to ask are: first, will the data tell you what you need to know? And second, how does the information integrate into users' decision-making routines? In transit, digital information provided by agencies was not presented in formats that met the mobility needs of riders. Further, static timetable information often did not reflect the actual arrival times of buses and trains. Once third-party developers could customize the information for mobile access and riders could consult the real-time location and arrivals within a transit system, transit system information improved in salience and in its ability to help users make better travel decisions on a daily basis.

Lesson 2: Prioritize the disclosure of data for which there is public demand. Along with disclosing schedules, routes and arrivals predictions, some transit agencies have also disclosed performance data. Yet, in this study, we found very little public use of this information. Public demand centered on information that would help transit riders make individual decisions about their own transportation needs at a given place and time. Knowing on-time-performance metrics for a particular bus route would tell riders, for instance, that they can expect their bus to be on time 62 percent of the time. That is a fundamentally different type of information than knowing that the bus you need to take at 8am on Monday will arrive in six minutes. The latter is actionable information whereas the latter is not.

What other types of information exhibit levels of demand and immediate salience similar to transit data? A very successful case of the public sector disclosure of information for third-party use is weather data. In Boston, Josh Robin and Chris Dempsey made their case for opening up MBTA data by using the National Weather Service as an example how public data releases can leverage third-party ingenuity to disseminate and embed essential information broadly. Like transit data, many people consult third-party sources for weather information on a daily basis. An analogous type of information and use case could be Health Department restaurant ratings, which help diners assess the public health risks of restaurants by providing timely and salient information through letter grades posted at restaurant entrances.[47]

Lesson 3: Determine whether information intermediaries play a role in the disclosure ecosystem and support the development of that ecosystem. When disclosure involves complex datasets, information intermediaries are likely to be the target audience of disclosure. In transit, these intermediaries were communities of software developers who were eager to use their skills to translate data into information that was easy for transit riders to access and understand through user-friendly applications. Disclosers should ask: who are the people or organizations interested and ready to translate data into formats and platforms that can reach end-users? These people and groups should be involved in the process of identifying the core public data to be disclosed and in defining the standards of quality and timeliness that are most valuable for users.

Like with digital signs displaying arrival times at bus stops and train stations, restaurant hygiene letter grades posted at restaurant entrances provide salient public health information after someone has already made an effort to arrive at a restaurant. For pre-dining planning, New York City's Department of Health and Mental Hygiene offers a searchable database of restaurant grades on its website, and built an iPhone and iPad mobile application for the public access the same information while on the go.[48] Nevertheless, third-party developers have also contributed to the City's efforts by customizing applications for other mobile platforms, like Android, and integrating the information onto other multi-purpose platforms like Foursquare. Also, the *New York Times* built an interactive map with the Health Department's database of restaurant inspections to accompany its reporting on common infractions by restaurants with "A" grades.[49] Like in transit, intermediaries of information also play a role in more broadly disseminating restaurant hygiene grades to prospective diners.

---

[47] See Fung, Graham and Weil's (2007) case study on Pg. 193-194. Also at http://www.transparencypolicy.net/restaurant-hygiene.php

[48] New York City Department of Health and Mental Hygiene Restaurant Inspection Information website http://www.nyc.gov/html/doh/html/rii/index.shtml

[49] White, Jeremy (2012)

Lesson 4: Adopt an open, non-proprietary data standard.
An open approach to data disclosure contributes to the sustainability of transparency systems. Open data standards and open source programming projects provide a foundation that enable data disclosers and users alike to learn from each other and adapt and customize existing tools in collaborative and iterative processes. Open sourcing projects also enable communities of developers to improve and maintain applications collaboratively, helping to insure against valuable projects dying due to their original coders losing interest. Portland's TriMet, for example, explicitly asks that third-party applications be open sourced so that in case a developer decides to leave a programming project behind, other developers or the agency itself can become maintainers of the project, avoiding situations where end users find broken or out-of-date applications.

Further, open data standards ensure that organizations retain the legal right to disclose their data and that information is interoperable with other datasets. They also allow third-party users of the information to reshape, remix and mash the data with other information to add to its functionality and value, depending on different use cases for different customers. Following in the steps of the GTFS transit data standard, New York City is collaboratively developing an open data specification for health inspections, paving the way for broader dissemination of restaurant hygiene information through third-party intermediaries.[50]

[50] New York City Open Data Tech Standards, nycopendata.pediacities.com/wiki/index.php/Restaurant_Inspection_Open_Data_Specification

References

"MBTA: App Judgment." September 5, 2009.*The Boston Globe*.

"Restaurant Inspection Information." New York City Department of Health and Mental Hygiene, accessed June 14, 2012, , http://www.nyc.gov/html/doh/html/rii/index.shtml.

American Public Transportation Association. 2012. *Transit Ridership Report: First Quarter 2012*. http://www.apta.com/resources/statistics/Documents/Ridership/2012-q1-ridership-APTA.pdf.

Annear, Steve. May 30, 2012. *MBTA Sets Longest Consecutive Ridership Record with 15 Months of Growth*. Boston Innovation. http://bostinno.com/2012/05/30/mbta-sets-longest-consecutive-ridership-record-with-15-months-of-growth-weekday-avg-tops-1-3-million-trips/.

Antrim, Aaron. 2008. *TriMet Innovations in Transit Data Publishing*. More Riders. http://www.trilliumtransit.com/files/Trimet%20data%20publishing%20interview.pdf.

Badger, Emily. April 23, 2012. *Why is a Patent Troll in Luxembourg Suing U.S. Public Transit Agencies?*. http://www.theatlanticcities.com/technology/2012/04/why-patent-troll-luxemburg-suing-us-public-transit-agencies/1819/.

Brunelle, Joey. 2011. *How's Metro?*. http://howsmetro.com.

———. 2011. *How's the T?*. http://howsthet.com/.

Fung, Archon, Mary Graham, and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency*. New York: Cambridge University Press.

Grynbaum, Michael M. September 27, 2009. "M.T.A. is Easing its Strict, Sometimes Combative, Approach to Outside Web Developers." *The New York Times*. http://www.nytimes.com/2009/09/28/nyregion/28mta.html.

Haddon, Heather. August 24, 2009. *MTA Looks for a Cut from Mobile Phone Applications*. AM New York. http://www.amny.com/urbanite-1.812039/mta-looks-for-a-cut-from-mobile-phone-applications-1.1392025.

Juliano, Michael. February 21, 2011. *Interview with Harper Reed*. A.V. Club Chicago. http://www.avclub.com/chicago/articles/harper-reed,51892/.

Kabak, Benjamin. January 11, 2012. *Real-Time Bus Tracking Comes to Staten Island*. Second Avenue Sagas. http://secondavenuesagas.com/2012/01/11/real-time-bus-tracking-comes-to-staten-island/.

Lathrop, David and Laurel Ruma, eds. 2010. *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Layton, Lyndsey. April 7, 2001. "Metro's Signs Sending the Wrong Message; Erroneous Reports Irk Riders, Workers." *The Washington Post*, B01.

McHugh, Tim. 2008. *Leveraging Resources for Customer Information by Exposing Transit Data*. APTA Trans|Tech 2008. TriMet. http://www.slideshare.net/headway/portland-trimet-presentation-on-data-sharing-updated-version.

Metropolitan Transportation Authority. June 2012. *Finance Committee Meeting*. http://mta.info/mta/news/books/pdf/120625_1245_Finance.pdf

New York City Open Data Tech Standards. "Restaurant Inspection Open Data Specification." New York City Department of Information Technology & Telecommunications, http://nycopendata.pediacities.com/wiki/index.php/Restaurant_Inspection_Open_Data_Specification.

O'Reilly, Tim. 2010. "Government as a Platform." In *Open Government: Collaborationn, Transparency, and Participation in Practice*, 11. Sebastopol, CA: O'Reilly.

Oshiro, Dana. August 20, 2009. *NY Transportation Authority Cites Schedules as Copyrighted Material*. Read Write Web. http://www.readwriteweb.com/archives/ny_transportation_authority_cites_schedules_as_cop.php#comments.

Perkins, Michael. March 23, 2009. *It's here! Metro Posts Transit Data Online*. Greater Greater Washington. http://greatergreaterwashington.org/post/1845/its-here-metro-posts-transit-data-online/.

———. August 6, 2009. *Metro Responds to Google Transit Posts*. Greater Greater Washington. http://greatergreaterwashington.org/post/3145/metro-responds-to-google-transit-posts/.

Raschke, Kurt. April 12, 2011. *Open standards for AVL and other real-time transit data*. Raschke on Transport. http://transport.kurtraschke.com/2011/04/open-standards-for-avl.

———. April 15, 2012. *WMATA's Half-Hearted Open Data Hurts Everyone*. Raschke on Transport. http://transport.kurtraschke.com/2012/04/wmata-eles-api.

Reed, Harper. October 6, 2008. *Chicago Transit API*. Harper Reed: Technology and Death Metal. www.nata2.org/2008/10/06/chicago-transit-api/.

Rogers, Everett M. 2003 (1962). *Diffusion of Innovations*. 5th ed. New York: Free Press.

Rosiak, Luke. December 9, 2010. "30% of Buses 6+ Minutes Off Schedule." *The Washington Post*. http://voices.washingtonpost.com/dr-gridlock/2010/12/metro_releases_bus_data.html.

Ross, Matthew. January 5, 2010. *How Google and Portland's TriMet Set the Standard for Open Transit Data*. SF.StreetsBlog.Org. http://sf.streetsblog.org/2010/01/05/how-google-and-portlands-trimet-set-the-standard-for-open-transit-data/.

Sanborn, Adrian and Chioma Madubata. 2011. *Where's that Bus??*. http://www.cs171.org/2011/projects/web/Sanborn-Madupata/applet/.

Shirky, Clay. March 30, 2004. *Situated Software*. Clay Shirky's Writings about the Internet. http://shirky.com/writings/situated_software.html.

Snyder, Tanya. 2012. *Patent Troll Sues Transit Agencies for Releasing Real-Time Transit Info*. DC.StreetsBlog.Org. http://dc.streetsblog.org/2012/04/16/patent-troll-sues-transit-agencies-for-releasing-real-time-transit-info/.

Tang, Lei and Piyushimita (Vonu) Thakuriah. 2012. "Ridership Effects of Real-Time Bus Information System: A Case Study in the City of Chicago." *Transportation Research Part C: Emerging Technologies* 22: 146-161.

Transit Cooperative Research Program. 2003. *Real-Time Bus Arrival Information Systems*. Washington D.C.: Transportation Research Board.

von Hippel, Eric. 2005. *Democratizing Innovation*. Cambridge, MA: The MIT Press.

von Hippel, Eric and von Krogh Georg. 2003. "Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science." *Organization Science* 14 (2): 209.

Washington Metropolitan Area Transit Authority. April 8, 2010. *Customer Services, Operations, and Safety Committee: Next Bus Performance*. http://www.wmata.com/about_metro/board_of_directors/board_docs/040810_NextBus.pdf.

Washington Metropolitan Area Transit Authority. May 2012. *Vital Signs Report: A Scorecard of Metro's Key Performance Indicators 2012 1st Quarter Results*. http://www.wmata.com/about_metro/docs/Vital_Signs_%20May%20QTRLY%202012.pdf

Watkins, Kari Edison, Brian Ferris, Alan Borning, G. Scott Rutherford, and David Layton. 2011. "Where is My Bus? Impact of Mobile Real-Time Information on the Perceived and Actual Wait Time of Transit Riders." *Transportation Research Part A: Policy and Practice* 45 (8): 839-848.

Weil, David, Archon Fung, Mary Graham, and Elena Fagotto. 2006. "The Effectiveness of Regulatory Disclosure Policies." *Journal of Policy Analysis and Management* 25 (1).

White, Jeremy. April 23, 2012. *New York Health Department Restaurant Ratings Map* The New York Times. http://www.nytimes.com/interactive/dining/new-york-health-department-restaurant-ratings-map.html.

Appendix I: Terms

API –API refers to an Application Programming Interface, which is a data access protocol system that allows different websites or applications to communicate with each other and extract or exchange information. APIs are necessary for data that changes quickly, like real-time arrival data. With static data like transit schedules and routes, bulk downloads work well.

App – The word "app" is shorthand for "application." An application is a piece of software.

Apps contest – Apps contests are events that invite software developers to innovate new applications from existing data sets. These contests feature prizes for best applications built. For example, the City of New York has hosted three NYC BigApps competitions.

AVL – Automatic Vehicle Location systems use hardware like GPS units and software to track the trains and buses in a transit network.

Countdown clocks/signs -- Electronic signs in subway stations or bus stops that let riders know, in minutes, when the next train or bus is scheduled to arrive.

Creative Commons license – a standardized copyright license that allows content to be copied, distributed, edited, remixed and built upon.

GPS – Global Positioning System, used by transit systems for above-ground vehicle tracking.

GTFS –General Transit Feed Specification. GTFS is an open, non-proprietary standard for transit schedules and route data that allows agencies to share data with each other and with the public.

GTFS-realtime – Google's data standard for real time transit data developed as an extension of the GTFS standard for schedules and routes. Additional data in GTFS-realtime includes vehicle positions, trip updates and service alerts.

Hackathon – A hackathon is an event where computer programmers get together to build software projects collaboratively over a short period of time, usually a weekend.

Open data – The term "open data" refers to data that is free for anyone to use, reuse, and redistribute, at most with an attribution or share-alike requirement (see Creative Commons license).

Open source – The term "open source" refers to both a type of software and a software development process. Open source software is built from computer code that is available for anyone to see, change, contribute to, and download. As a process, open source involves freely exchanging, modifying and building upon others' software both individually and collaboratively among computer programmers.[51]

Patent troll – "A term used for a person or company who enforces patents against one or more alleged infringers in a manner considered aggressive or opportunistic with no intention to manufacture or market the patented invention."[52]

Screen scraping – The process of using a computer program to extract data from websites to be analyzed or reused in applications.

Shapefiles – A type of data file that contains mapping information like the geographic coordinates of a bus route.

SIRI – Service Interface for Real Time Information. A open data standard for real time transit information that allows access to AVL data streams and service updates.

Source code – The list of commands written by computer programmers to build software systems.

Vendor – Vendors are private contractors that provide technical services and software to transit agencies.

---

[51] This is a composite definition building from von Hippel and von Krogh (2003).
[52] Quoted from Wikipedia's definition of a Patent Troll http://en.wikipedia.org/wiki/Patent_troll, accessed June 13, 2012 .

Appendix II: Methodology

Case Selection

The findings in this report are drawn from five case histories that examine the process by which some of the largest transit agencies in the U.S. have disclosed their operations data for public use. In order to select transit agencies for in-depth study, we began by placing the dates of transit data disclosure by agency on a timeline (see Figure 1 in above report). We noticed a distinct pattern whereby transit agencies first released static schedule data to the public, followed by real-time data on vehicle positions. We also noted a snowball effect revealing that the open data approach spread quickly across transit agencies in the U.S., resembling Everett Rogers' notion of how new ideas and technologies diffuse across organizations.[53]

Realizing that the experience of opening up transit data to customers fits the diffusion of innovations framework, we used Rogers' five innovation adopter categories as a guide in selecting the transit agencies for our case studies.[54] These categories are:

- Innovators: these are the first to adopt an innovation and are willing to take risks. They serve a critical role in influencing others.
- Early adopters: these are the second fastest adopters of an innovation and serve as role models known for "judicious innovation decisions".[55] They help trigger further adoption by others.
- Early majority: these organizations tend to be slower in adopting new ideas because they tend to deliberate for some time before completely jumping on board with an innovation.
- Late majority: these adopt innovations with skepticism and caution, and ultimately adopt and innovation as a result of increasing peer pressure.
- Late adopter: the most tradition-bound, late adopters need to be certain that a new idea will not fail. They tend to engage in lengthy innovation-decision processes.

Each transit agency selected for this study corresponds loosely with one of the adopter categories defined above (see Figure 1 and Table 1 in above report). In the cases where more than one transit agency fit an adopter category – as with San Francisco's MUNI and Boston's MBTA – we selected the latter due to our greater access to agency officials and other relevant sources such as local software developers and transit customers in Boston. Further, Boston as an "early adopter" and Chicago as an "early majority" case closely follow each other in timing their data releases to the public, we chose to categorize the Chicago Transit Authority (CTA) as an early majority agency due to their mixed

[53] Rogers, Everett. 2003 (1962). *Diffusion of Innovations*. New York: Free Press.
[54] Rogers (2003), Pg. 281.
[55] Rogers (2003), Pg. 283.

strategy in creating customer-facing information interfaces in-house first, and later releasing raw data-feeds to software developers. The MBTA approach, in contrast, is a more pure open data strategy where the agency recognized its limited capacity to build dynamic customer-information systems and therefore actively fostered the co-production of these services by courting the local software developer community before opening up data from their operations systems.

Case Studies

Each case study is based on semi-structured interviews with agency officials, developers of transit applications, and other sector experts. Given that there was variation within transit agencies as to where the momentum was in implementing the disclosure of operations information, we interviewed transit agency officials working in all aspects of transit management: the executive office, information technology, operations technology, customer information, and systems engineering. To understand the motivations of users of this data, we identified software developers who were working with operations data to build transit apps in each case study city. We identified developers through transit agency officials and, where available, we searched the transit agencies' App Showcase web pages[56] to identify apps developers. Other sources for identifying apps developers included Google groups focused on transit application development and online blogs that track advances in transit and civic technologies.[57] We supplemented this approach by searching the Apple and Android apps stores for other relevant transit applications.

When interviewing transit officials, we discussed the institutional factors that led to implementing an open data approach within that agency, noting both the motivations and challenges involved in the process. We also noted how agencies manage their releases of data, ensure data quality, and establish relationships with local software developers. In speaking with developers, we principally asked about their motivation for using transit data and their experience in working with transit agencies. We were also curious as to which considerations drove the design of the applications they built with transit information and the feedback they have received from users of their apps. Where available, we also recorded the approximate number of downloads of the apps built by the developers we interviewed.

These interviews yielded a rich story of the process by which transit agencies disclosed their operations data to the public and the motivations that drove the developer community to work with complex temporal and spatial data. The case studies reveal this story and are the focus of this report.

---

[56] See the MBTA's App Showcase at http://www.mbta.com/rider_tools/apps/, the CTA's App Center at http://www.transitchicago.com/apps/, and TriMet's App Center at http://trimet.org/apps/index.htm.
[57] There are several Google Groups established to facilitate communications between software developers interested in working with transit data, including the Transit Developers group (http://groups.google.com/group/transit-developers).

But since we are also interested in understanding the effects of transit data disclosure on transit riders, our research team has sought a quantitative approach to capturing the public's response to dynamic information about bus and train arrivals by surveying transit riders. The Transparency Policy Project is collaborating with Candace Brakewood, PhD candidate at MIT's Engineering Systems Division, and Professor Chris Zegras of MIT's Department of Urban Studies and Planning on these two streams of research. While these analyses are still in progress, they nevertheless inform the case study findings discussed in this report, and to that end, we discuss the methodology for this aspect of our research below. Findings from this quantitative analysis of transit data rider effects are forthcoming.

Customer Surveys

As a means to capture changes in rider behavior and attitude about transit service, we are fielded on-board customer surveys for two of our case study transit systems: New York's MTA and Boston's MBTA. We employed pre- and post-wave surveys where we asked riders about their transit behavior and attitudes regarding service both before and after agencies made real time transit data made available to the public. The focus of our surveys in New York is to understand the effects of real time information on riders who rely on bus service, whereas the focus of our work in Boston is on the effects of real time information on riders of commuter rail.

We partnered with New York's MTA to design and distribute surveys for the B63 bus route in Brooklyn, the first in New York to offer real time arrival information to riders through the MTA's BusTime pilot program.[58] We collected 271 surveys during January 2011 before the MTA launched BusTime, which makes real time transit information available to riders of the B63 bus. In June 2010, four months after introducing real time arrival information on that route, the MTA conducted a follow-up survey on the B63 line, yielding 762 responses. As part of this effort, the MTA also included a control route along the B62 bus route (which does not provide real time arrivals to riders) as a way to measure changes in rider attitudes and behavior attributable to BusTime information, collecting 567 surveys. The MTA is following this same survey approach to capture rider effects from the Staten Island-wide rollout of BusTime for every bus in that borough with a pre-wave survey in January 2012 and a post-wave survey in May 2012. Our research team continues to collaborate with MTA on this research.

The methodology in Boston also involved a pre- and post-wave survey of transit riders, but instead focused on the MBTA's commuter rail service. Prior to the agency's release of real time data for its commuter rail system in June 2011, we distributed on-board surveys to riders on two lines, the Worcester line out of South Station and the Newbury/Rockport line out of North Station. Both are high-ridership lines with approximately 17,000 weekday boardings, but in 2010

[58] For more on the MTA's BusTime pilot project, see http://bustime.mta.info/wiki/About

they diverged in their on-time performance measures (defined as arriving five minutes or less after the scheduled time). We collected 881 completed surveys that capture current traveler characteristics, including utilization of transit information sources, and levels of customer satisfaction. We distributed the post-wave survey in June 2012 to assess the extent to which the provision of real time arrival information influenced baseline measures of information use, travel behavior and customer satisfaction among commuter rail riders.

Through these on-board customer surveys, we seek to produce statistically significant results to the question of whether access to real-time operations information on buses and commuter rail produces changes in transit riders' travel behavior and level of satisfaction with service.

Appendix III: Case Studies

The Innovator: Portland's TriMet

Tri-County Metropolitan Transportation System of Oregon

Portland's Tri-County Metropolitan Transportation System of Oregon (TriMet) was the first transit agency in the United States to release machine-readable schedules and real-time operations data to the public. TriMet had installed GPS-enabled vehicle location systems on its buses in the mid-1990s and by 2003 the agency's website allowed customers to query transit lines and receive estimated arrival times. This online tool complemented LED signs at transit stations that also displayed arrival times for streetcars and buses. These were dynamic information systems, built and controlled by TriMet. Three years later, the agency's customer information strategy evolved towards a more open and collaborative one when it exposed its transit data to the public. As a result, TriMet's timetable data files were the first to be accessible on Google Maps and used by Portland's software developer community to build transit applications independently of the agency.

The decision to publicly release TriMet schedules in a standardized and open electronic format came about through a combination of forces. First, forward thinking engineers within the agency saw the value of leveraging multiple sources of information to help riders make the best transit choices. Second, the agency had the in-house information technology infrastructure and capacity required to generate and expose its own transit data. Meanwhile, do-it-yourself technologists outside the agency were already scraping data from the TriMet website to customize the information for their own commutes. Once bus and streetcar schedules were up on Google Maps (now known as Google Transit) in the GTFS format, TriMet reasoned that as a public agency, they should not be exclusive to just one party with their data. So the GTFS files also went up on the TriMet website for download. This approach offered interested software developers a more stable, timely and reliable dataset than what they got from screen scraping data from the agency's website. It also guaranteed that independent developers would abide by TriMet's terms of use, insulating the agency from any potential legal risk in case customer-facing applications were to misrepresent the transit data.

The resourcefulness of a few individuals sparked the process by which TriMet exposed its data to the public, and in turn, influenced the path of other transit agencies to follow TriMet's lead. While traveling abroad in 2005, Bibiana McHugh, TriMet's IT Manager of Geographic Information Systems, thought that transit routes and schedules should be as easily available online as driving directions. She approached Google, Mapquest and Yahoo with her idea, offering TriMet as a partner to add transit information onto existing online maps. McHugh received a response from Chris Harrelson at Google, who had already been working on integrating transit data into Google Maps as his "20 percent time" project.[59] Since TriMet had a centralized database for its operations information, McHugh and her husband Tim (the agency's Chief Technology Officer) were able to quickly pull the necessary temporal and spatial data elements, such as route numbers, route shape files, stop locations, and stop times. TriMet's timetable data became available through Google Maps on December 2005. TriMet and Google then worked together over the following nine months to develop a data standard for transit schedules, which became the General Transit Feed Specification (GTFS).[60] By the end of 2006, TriMet had made their GTFS file public under a Creative Commons license.[61]

TriMet was able to move quickly in exposing its data to the public because it owned and managed its own information technology resources and, consequently, did not rely on outside vendors to generate operations data for its transit system. Institutionally, TriMet was well positioned to see the value of, and act upon, the notion of "creating more opportunities for easy access to transit information and for exposure to it."[62] Its information technology, communications and marketing functions were all housed within the agency's Communication Technology department, likely facilitating decision-making about such an innovative approach to delivering customer information. Further, TriMet employed in-house developers and therefore had the capacity to build their own services and tools, including a real-time arrival prediction algorithm, which precluded the need to contract with a vendor like Nextbus for real time feeds of transit information. With an understanding of how information technologies and customer information were coming together, and without outside licensing constraints on its operations data, TriMet was able to pursue an open data strategy.

In a city as focused on civic innovation as Portland,[63] citizens also played a role in motivating TriMet to expose their operations data. Chris Smith was a local software developer and transit advocate with the group Portland Transport. In 2006 he started screen scraping TriMet's website for real-time arrival information to build a mobile application to improve his own commute. Three different transit lines whose stops were a few blocks apart linked his neighborhood to his job downtown. Smith built the Transit Surfer application so that he could compare arrival times for the three bus lines on his smart phone while on the go.[64] Smith's approach customized and improved upon TriMet's TransitTracker web service, which at that point offered information for one transit stop per query and was optimized for websites, not mobile phones. He showed Transit Surfer to people he knew at TriMet, pitching for an open API to facilitate other third-party apps. Eventually TriMet asked Smith to begin experimenting with a

---

[59] Ross, Matthew (2010).
[60] The GTFS data standard was originally called the Google Transit Feed Specification.
[61] A standardized, open copyright license that allows sharing of content. See http://creativecommons.org/licenses/
[62] Tim McHugh quoted by Antrim, Aaron (2008).
[63] Portland held the first Civic Apps contest in early 2010. See http://www.civicapps.org/press/city-portland-launches-civicapps-design-contest
[64] See Chris Smith's list of favorite trips at http://tsrf.us/cps.html. Built for the earliest generation of smart phones in 2006, the information display is in HTML text only (the iPhone was released a year later, in 2007).

prototype API. Later Smith learned that Bibiana McHugh had been lobbying from within the organization to expose the operations data to the public as well. Smith remarked: "Bibiana was pushing to get the data out and I came along pushing the door to get in and we broke through!"[65]

The agency's concerns about exposing operations data to the public centered on mitigating any legal liabilities that could arise from third-parties using TriMet's data. Bibiana McHugh worked closely with the agency's legal department in setting the Terms of Use for the data. Counter to conventional thinking among transit agencies (see the WMATA and MTA cases), TriMet realized that an open data strategy could actually insulate the agency from legal risk. As stated in the Terms of Use, developers were asked to agree that TriMet made the data available "as is" and "as available" and that:

> you use the Service and Data at your own risk, and you assume the risk that the Service or Data may provide incorrect information to you or your workers, as well as the risk that any material downloaded by you from the Service may cause loss or data or damage to your computer system.[66]

TriMet also needed a policy on featuring third-party applications built with agency data on their website. After much deliberation, TriMet decided that third-party applications could be included on its "App Center"[67] webpage if they met two simple criteria:

1. The apps used TriMet's data resources (meaning, they did not scrape the data off the agency website); and,
2. The apps worked the way they said they would.

TriMet did not regard its role as being one of deciding whether an application worked well or not – as long as it performed the function it advertised, then an application could be added to the agency's website. The agency believed that if customers did not like the way certain apps worked, then they would not be popular apps. Feedback from customers directly to software developers would eventually improve applications' functionalities. As Bibiana McHugh commented regarding TriMet's role in the ecosystem of transit transparency: "Our responsibility ended with making the data available."[68]

Interviews with software developers who built transit applications using TriMet data indicate that the agency took on a broader responsibility than "just making the data available."[69] Bibiana McHugh was a critical conduit of communication between the developer community and her organization. Issues included troubleshooting data bugs, answering questions, and understanding developer

---

[65] Author interview with Chris Smith.
[66] TriMet Terms of Use http://developer.trimet.org/terms_of_use.shtml
[67] TriMet App Center http://trimet.org/apps/
[68] Author interview with Bibiana McHugh.
[69] Author interview with John McBride.

needs. “Bibiana and her people were terrific. They did everything they could to help me,”[70] said John McBride, a developer who built myTrimet.com in 2007. Andrew Wallace, the developer behind the PDX Bus iPhone app, concurred: “When bugs came with the data, I contacted TriMet and they fixed them quickly.”[71] Chris Smith, who built the Transit Surfer application, agreed, “TriMet [was] very responsive to developer input and requests.”[72] Interactions between developers and McHugh happened both online and offline. An online developer discussion group served as an important platform for questions, feedback, troubleshooting and announcements. Just as important were McHugh’s efforts at community building: she would meet for lunch with local technologists who used transit data and in turn, developers knew they had a personal contact inside the agency. Trust between the agency and the developer community facilitated an iterative refinement of data and clarification of user needs.

By offering public access to its GTFS schedule files and real-time data feeds, TriMet was able to benefit from the efforts of a community of software developers who, to date, have built 45 different customer-information tools that vary in format, content and technical specifications, meeting multiple user needs. By building an API, TriMet also avoided “addressing multiple and varied data requests on an individual basis,” saving engineers time and resources.[73] And by allowing third-parties to expand the sources for transit information for customers, the agency furthered its commitment to helping commuters chose transit over single occupancy vehicles.

As this research shows, the open data strategy pioneered by Portland’s TriMet spread quickly to other transit agencies in the United States, but with varying degrees of success.

[70] Author interview with John McBride.
[71] Author interview with Andrew Wallace.
[72] Author interview with Chris Smith.
[73] McHugh, Tim (2008).

The Early Adopter: Boston's MBTA

Massachusetts Bay Transportation Authority

Boston's Massachusetts Bay Transportation Authority (MBTA) was an early adopter of the open data approach pioneered by TriMet in Portland. But three years passed between when TriMet exposed its operations data to the public in 2006 and when the MBTA began disclosing its own transit data in the fall of 2009. In those years, the MBTA had invested in upgrading its website's trip planning software and had begun planning a real time arrival information system for customers. The MBTA's planned bus arrival information system relied on installing digital countdown signs at bus stops, and riders accessing arrival information on the MBTA website, via a call-in service and with mobile text messaging. But this was to be a closed system, managed and executed entirely in-house by the MBTA and its vendor contracts. The story of how the MBTA shifted course and reversed its customer information strategy – releasing data first, and installing infrastructure later – is centered around two young political appointees at the Massachusetts Department of Transportation (MassDOT) who were inspired by Portland's vibrant ecosystem of independent, third-party transit applications.

Chris Dempsey was Deputy Chief of Staff and Joshua Robin was Manager for Performance Reporting in the Massachusetts Executive Office of Transportation, which later became MassDOT and incorporated the MBTA. In the spring of 2009, Dempsey was working on the 511 real time highway traffic information system supported by the Federal Highway Administration. He saw that other sources of information – particularly local websites, radio, and television – were better at disseminating traffic information than the 511 system, which they estimated reached only one percent of commuters. Robin, meanwhile, was responsible for producing the MBTA's performance scorecard and worked closely with engineers in the agency's operations technology (OpsTech) department to produce monthly reports of performance metrics for the transit system.[74] Robin saw that the OpsTech engineers had the ability to access real time information on all buses in the city at the agency's control center, on their desktop computers, and on their mobile devices. "Why can't the public have access to this information too?" asked Robin.[75]

Looking to TriMet's experience, Dempsey and Robin believed that a public agency like the MBTA should be an "information wholesaler," providing public data that people with technical skills can then transform into various applications, targeting different user needs, and at no cost to the agency. In arguing their position, they used the National Weather Service as an analogy: it provided weather forecasts in a format that enabled a wide array of third-parties to present weather information.[76] With a concentration of tech-startups, university students, and other technologists in the Boston region capable of

[74] See the MBTA's monthly ScoreCard at http://www.mbta.com/about_the_mbta/scorecard/
[75] Author interview with Joshua Robin.
[76] Chris Dempsey discusses the National Weather Service model in " A Case for Open Data in Transit" by StreetFilms http://www.youtube.com/watch?v=kf2Kx5HhkGg&feature=related

working with complex datasets, they believed local transit riders would benefit from a similar approach to accessing transit information.

In the summer of 2009, Dempsey and Robin saw a window of opportunity to push for open data as MassDOT and the MBTA were restructured in preparation for new leadership. The temporary power vacuum allowed the young appointees enough latitude to act on their vision. It also helped that the MBTA had just finished compiling a GTFS file for Google Transit after struggling to overcome obstacles around proprietary licenses, data formats and institutional resistance. It took four years to produce a reliable GTFS file for such a large transit network. On a daily basis, the system handled 1.1 million trips with 500,000 opportunities for riders to board and exit a bus. Some of the vendors that provided software for the MBTA to generate its scheduling and routing information did not allow the agency to distribute this data to third-parties. Within the organization itself, engineers questioned whether expanding the sources of transit information beyond the MBTA's control would compete with their website's trip planner, which they were working on upgrading. There was also concern about third-parties delivering inaccurate information to customers. Once MBTA overcame these technical and institutional barriers to joining Google Transit, Dempsey and Robin pushed the agency further, asking them to provide the GTFS file to the public as well, as TriMet had done years earlier.

Hoping to nudge the agency towards an open data strategy by producing a successful "proof of concept," Dempsey and Robin formed the MassDOT's Developer's Initiative.[77] With the help of Laurel Ruma, then the Government 2.0 evangelist at O'Reilly Media,[78] they organized an event in early August 2009 to engage local software developers in a discussion about the opportunities and challenges of implementing open data for transit. According to Ruma, about 25 people attended the first meeting. Participants were primarily software developers who worked at tech companies in the area. The atmosphere was combative: *why can't we have the data now? You must know where every train and bus is! This is public data, paid for with public money, why can't we have it?* There was an expectation from the developer community that this data already existed within MassDOT and that it should be made available as soon as possible.

To get the developers involved, Dempsey and Robin understood that they urgently needed *some* data for Boston's transit system, even if it was not absolutely perfect or complete. As Robin stressed to his colleagues at the MBTA: "let's focus first on the data, it's more important and a better value."[79] Within a week of the developer meeting, they had convinced engineers in the operations technology department of the MBTA to post schedules online in GTFS format. This first data release quickly yielded a couple of smartphone apps and a homemade LED countdown sign installed in a coffee shop – at no cost to the agency. These were valuable proofs of concept that helped Dempsey and Robin

[77] MassDOT Developers Initiative page http://www.eot.state.ma.us/developers/
[78] O'Reilly Media is based in Portland, Oregon, but Ruma worked from Boston. Tim O'Reilly, the founder of O'Reilly Media is an open government advocate and coined the term "government as a platform." See O'Reilly's chapter "Government As a Platform" in Lathrop and Ruma (2010)
[79] Author interview with Joshua Robin.

gain support from the state's Secretary of Transportation. Plus it earned the MBTA a positive editorial from the *Boston Globe*, a rare accolade for the transit agency.[80]

The timing of MassDOT's efforts to expose operations data was opportune. The MBTA and Nextbus already had a pilot program in place to provide real time bus information, although the strategy relied on installing expensive countdown signs at bus stops and managing a call and text messaging service. MBTA engineers hesitated to open up the Nextbus feed to the public at such a preliminary stage of the project – it only covered five of the 200 bus lines in the city. But Dempsey and Robin realized that piloting transit data in concert with the developer community allowed local technologists and the MBTA to collaboratively build up skills, troubleshoot problems, and ultimately improve data feeds over time. By late fall, MassDOT and the MBTA announced the availability of real time prediction data for those five bus routes at its first developer's conference.[81] As David Barker, the Director of the MBTA's Operations Technology department recalled "We got a great response. By the time we came back from the data launch event, there were dots moving on Google Earth."[82] Indeed, an hour after the data release, a developer had built an application displaying bus locations on a map in real time.

Strong support for open data from the heads of MassDOT, the MBTA, and even Governor Deval Patrick gave the agency momentum to release more operations data. High levels of enthusiasm and commitment from the developer community generated dozens of different ways for riders to access transit information. Within a year, the MBTA had released scheduling and real time data for the entire MBTA system, with the exception of the Green Line trolley and Commuter Rail (eventually released in July 2011). MBTA engineers like Barker learned that "putting the data out there – letting others do the work on our behalf, turned out to be more cost-effective."[83] The MBTA leveraged a one-year, $100,000 Nextbus contract for real time arrival predictions data and an API, into over 35 different customer-facing applications built by independent software developers.[84] Still, the MBTA did not abandon its plan to install digital signage as another way for riders to get real time arrival information – it just scaled it back. Instead of having countdown signs at every bus stop, where installation and maintenance are expensive, the agency placed LCD screens showing arrival predictions at high-traffic, intermodal stations where subways, buses and regional trains converge. Moving forward on its strategy to deliver better customer information (and in payment systems as well), the MBTA is looking for ways to support

---

[80] "MBTA App Judgment," *The Boston Globe.* September 5, 2009.
[81] The event agenda is posted online http://massdotdevconference-eorg.eventbrite.com/.
[82] Author interview with David Barker.
[83] Author interview with David Barker.
[84] Comment by Richard Davey, MBTA General Manager (2010-2011), at Harvard's Kennedy School of Government during a presentation titled "Public Transit in Tough Fiscal Times." March 24, 2011. From author's notes.

digital technologies, particularly mobile phones, which allow customers "to bring their own infrastructure."[85]

Like in Portland, critical to the high level of data take up in Boston was Dempsey and Robin's dedication to fostering face-to-face and online relationships with local software developers. The timeline below (Table A1) shows that MassDOT invited developers to monthly meetings and special events, like conferences and hackathons, as it ramped up its efforts to make public the MBTA's operations data in late 2009 and early 2010. Progress meetings complemented data releases as a way to check in with developer needs, clarify data formats, manage expectations and anticipate future releases. Developers were also instrumental in beta testing data feeds for the MBTA before they were released more broadly to the public.

These interactions with the agency gave developers confidence in the collaborative process within which they played an integral role. George Schneeloch, an Android app developer, valued the face-to-face meetings because "it was a good way to learn about the direction the MBTA was heading," and provided a rich opportunity for peer learning that complemented the online discussion group.[86] Dan Choi, who built the OpenMBTA application, felt "It was reassuring for MassDOT to be there in person. Their effort is obviously not an afterthought – they are consciously supportive of developers."[87] At an MBTA developer meeting held on the same night that the Boston Bruins were playing for the Stanley Cup in 2011, Richard Davey, General Manager of the MBTA, made a surprise visit to tell the 20 developers in attendance: "Thank you for coming and spending time with the team to make the T better. I appreciate you guys helping us a lot. It's your T, it's your system."[88]

As an early adopter of the open data approach pioneered by TriMet, the MBTA helped trigger other large agencies in the U.S. to consider releasing their data files to the public. Transit agencies in Chicago and New York looked to Boston for lessons on how to capture the benefits of this innovation and alleviate fears about exposing their operations data to the public. The establishment of personal networks by data champions like Bibiana McHugh, Joshua Robin and Chris Dempsey was critical to this diffusion of innovation process.

[85] Comment by Joshua Robin at Harvard Kennedy School event "Where's My Bus… and Beyond," November 3, 2011, sponsored by the Transparency Policy Project, Rappaport Institute for Greater Boston and the Taubman Center for State and Local Government. From author's notes.
[86] Author interview with George Schneeloch.
[87] Author interview with Dan Choi.
[88] General Manager Richard Davey, June 15, 2011. From author's notes.

Year	Date	Event
2009	August 4	First developers meeting
	August 14	MBTA schedule data (GTFS) released to the public
	September 15	Developers meeting
	September 28	MassDOT Developers Challenge announced
	October 20	Developers meeting
	October 29	OpenMBTA app by Dan Choi available on the Apple app store
	November 14	MassDOT Developers Conference at MIT Developers Challenge winners & trial real time feed announced
	December 8	MassDOT developers holiday party
2010	January 29-30	MassDOT/MBTA/MIT Hackathon
	February 18	MassDOT Developers Real Time Challenge announced
	March 24	MassDOT Real Time Challenge winners announced
	April 28	Progress meeting with developers
	June 3	"Where's the Bus 2.0: the wait is over" event. MBTA announces real time data available for all bus routes by the end of summer 2010
	June 14	Progress meeting with developers
	July 26	MBTA releases real time data for an additional 118 bus routes
	August 9	Progress meeting with developers
	September 7	MBTA launches apps gallery www.mbta.com/apps
	October 1	MBTA releases real time data for all subway lines except for green line
2011	June	MBTA real time data on Google Transit
	June 15	Developers meeting to announce Commuter Rail real time data feed
	June16	MBTA releases real time data feed for Commuter Rail system

Figure A1 Timeline of MassDOT/MBTA open data milestones and events (source: compiled from author's research and interview notes)

The Early Majority: Chicago's CTA

Chicago Transit Authority

The Chicago Transit Authority (CTA) launched its Bus Tracker program in 2006 by contracting with Clever Devices, a vendor that installed GPS units on buses and provided real time vehicle location tracking for the agency and its customers. The system was originally installed to comply with Americans with Disabilities Act requirements that buses integrate fully automated audio and visual announcements onboard. This automated vehicle location (AVL) system also allowed the CTA to better manage its bus fleet by reducing delays and bunching, and offered riders the ability to access estimated arrival times for bus routes on the CTA website.[89] The CTA had not anticipated exposing this real time data feed beyond its own website interface and Clever Devices had not planned on building an API for the data either. But in October 2008, Harper Reed, a local software developer and entrepreneur, built his own API for Bus Tracker. Reed wanted to help others build mobile applications to access the CTA's real time data feeds in a faster and more intuitive way than what was available on the agency's website. Reed documented what he had done on a wiki and wrote a blog post.[90] Almost immediately, his friends and other local developers began using Reed's API to build their own mobile apps for real time transit data in Chicago.

As an agency, the CTA was not overly protective of its data. It had joined Google Transit in the spring of 2008, at the time, the largest participating transit agency. It also quietly gave its schedule files to people who asked for them. Harper Reed approached the CTA together with Dan O'Neil, a civic-minded advocate and entrepreneur who knew the agency's General Manager Ron Huberman.[91] After this meeting with Reed and O'Neil, officials at the CTA reasoned they had three options for how to deal with Reed's API:

1. They could shut it down;
2. They could continue to let Reed and O'Neil run the API as an unofficial, unsupported and undocumented service; or,
3. They could legitimize the open data approach and build an official CTA-supported API for the real time bus data feeds.

To the CTA, Reed's API signaled demand for this data and so the agency was reluctant to shut it down. But if they were to allow the API to be unofficial and unsupported, the agency risked breaking all the applications that relied on the API if there was a change in the system that generated the data feeds. But internal deliberations on the third option, to legitimize a public API, centered on the question of whether it would be possible for the cash-strapped agency to

---

[89] See BusTracker on the CTA website http://ctabustracker.com/bustime/home.jsp
[90] Reed, Harper (2008).
[91] Before the real time arrivals data API, O'Neil and Reed had automated a twitter-based CTA alert service (@ctatweet) that combined crowdsourced alerts and official CTA alerts. O'Neil knew General Manager Ron Huberman through his brother Kevin O'Neil who runs the CTA Tattler blog http://www.chicagonow.com/cta-tattler, which reports on all things seen and heard on Chicago's transit system.

monetize the data in any way. The CTA looked to what other agencies had been doing in this regard, particularly the experience at TriMet, and reasoned that the data would be of greater value to customers than the small amount of profit the agency could potentially extract from licensing the data to third-party users. Plus, they had already given Google Transit their GTFS files for free. Tony Coppoletta, the CTA's Manager of External Electronic Communication, recalled that after hearing from Reed and O'Neil, and getting further encouragement from MassDOT's Chris Dempsey, the CTA decided to take an open approach to its real time data.

Working together with Clever Devices, the CTA made public its BusTracker API in November 2009. A CTA Developer's Center web page with data documentation followed in January 2010.[92] By January 2011, the CTA also offered a TrainTracker real time arrival information service for its elevated train network in Chicago.

In the year lag between the unofficial and the official API for CTA real time bus arrival information, developers had already built mobile applications using Reed's unofficial API. Frustrated with the CTA's BusTracker website because it only displayed arrival predictions according to bus numbers and did not work well on mobile phones, Chris Cieslak built the Buster iPhone app so he could consult his bus on the go and by route name ("Milwaukee" instead of "56"). After the CTA released its official API, the agency asked Cieslak and other developers to migrate their apps to the public API. Clever Devices, the CTA's vendor, even offered to purchase the Buster app, but Cieslak declined their offer.

Cieslak and others interviewed for this study remarked that the CTA had provided outstanding support for developers. The agency's documentation for their data and real time feeds was "Better than a lot of tech companies. It [was] actually easier to access CTA data than the twitter API."[93] Having Tony Coppoletta as a point person within the transit agency (like at TriMet with Bibiana McHugh and at MBTA with Joshua Robin) was also critical to facilitating third-party use of the agency's information. Coppoletta explained that the agency saw developers as "an extension of what our capabilities are and we want to be supportive of that."[94] Like TriMet and the MBTA, the CTA built an App Center to display applications that use Chicago transit data. Consistent with the other two agencies, the CTA wanted to support but not endorse third-party efforts. As long as developers abide by the agency's terms of service, are not broken or do anything malicious, they are included on the App Center.[95] To date, the CTAs App Center features 25 developer-built applications for transit data.

Among the five agencies we examined, Chicago has the broadest array of ways that transit data reaches people – extending beyond desktop and mobile apps to digital signage installed throughout the city. In line with the MBTA's notion that

---

92 The CTA's Developer Center http://www.transitchicago.com/developers/
93 Author interview with Chris Cieslak.
94 Author interview with Tony Coppoletta.
95 The CTA's App Center http://www.transitchicago.com/apps/

people are increasingly bringing their own physical infrastructure to the table, independent user communities like neighborhood associations, small businesses, and even apartment buildings used the CTA's API to display transit data on their own computer or LCD screens. This was facilitated by the CTA providing detailed instructions for a *Do-It-Yourself Transit Info Display* on its Developer Center webpage.[96] To support this service, the CTA generated a web-based slide show that included transit arrival information for nearby buses, service alerts, and local weather. Displays were installed in office building lobbies, apartment building entrances, coffee shops, and anywhere else people wanted to consult transit information.

The CTA's experience to exposing their operations data to public represents a hybrid model similar to TriMet. The agency began by building a core set of web-based, platform neutral customer information tools for real time bus arrivals. Along the way, the CTA learned that by expanding access to their operations data through a public API and data releases, local technologists could extend the agency's capabilities by building their own applications, meeting a broader variety of user needs.

[96] The CTA's DIY Transit Info Display page is at http://www.transitchicago.com/developers/diydisplay.aspx

The Late Majority: Washington's WMATA

Washington Metropolitan Area Transit Authority

At around the time TriMet was preparing to open its data to the public, Washington D.C.'s Metropolitan Area Transit Authority (WMATA) was positioned to lead the way in real-time customer information for its rail and bus systems. By 2001 WMATA had installed LED signs announcing the next arriving train in all 83 of its Metrorail stations, ahead of other agencies.[97] In 2007 WMATA had contracted with Nextbus to provide real time bus arrival information for internal performance management and customer information purposes. The plan was to offer riders bus arrival predictions on the WMATA website and through a call-in and text messaging service. But soon afterwards, unreliable data for bus predictions drove WMATA to cancel the Nextbus contract, just as smartphones were allowing people to access more information on the go. This marked the beginning of a contentious back and forth between advocates for open data in the Nation's Capital and a fiscally constrained public agency under pressure to expose its data.

Greater Greater Washington (GGW), a local-interest blog started in early 2008 by David Alpert, spearheaded the campaign to get WMATA's schedules on Google Transit and publish the data files online so software developers could also use the information to create transit applications. As Luke Rosiak, a local data journalist called them, GGW were a group of "young progressive urbanist digiphiles."[98] After disappointing responses from WMATA's customer service department, the bloggers put transit transparency squarely on the agency's radar when in December 2008 it delivered a petition with over 700 signatures (gathered in under a week) directly to the WMATA Board members' individual email accounts.

Greater Greater Washington was relentless in its coverage of WMATA's lack of transparency. In response to this pressure, WMATA posted a GTFS file of schedule and routing data for the public to download in late March 2009. But GGW considered the agency's terms of use to be overly restrictive, potentially requiring third-party developers to indemnify WMATA and raising the possibility that users of the data may have to pay a licensing fee in the future.[99] Plus, the agency's GTFS data was unreliable enough that Google would not end up signing an agreement with WMATA until July 2010 and WMATA data wouldn't appear on Google Transit until May 2011, over two years after they released the first GTFS file to the public.

One of WMATA's challenges in opening up its operations data to the public was a cautious institutional culture that during a severe fiscal crisis was forcing managers to focus on finding new ways to generate revenue. Another substantial obstacle was the agency's constrained internal capacity to resolve vexing data quality problems borne of overlapping legacy IT systems managed by different departments within the organization.

---

[97] Layton, Lyndsey (2001).
[98] Author interview with Luke Rosiak.
[99] Perkins, Michael (2009).

Similar to deliberations in Boston and Chicago, WMATA had invested in a online trip planning software that it believed offered comprehensive information to its customers. They did not consider enabling outside venues for accessing this information to be a priority and, as a further disincentive, they were receiving advertising revenue for web visits to their trip planner. The agency also believed it could further monetize the data and was reluctant to give it away for free to Google, in particular. In a response to GGW's queries about the lack of an agreement between WMATA and Google, the agency wrote:

> We plan to begin a study this year which will give us a firm idea as to commercial value of intellectual property like scheduling information. Until that study is completed it seems wisest to avoid entering into an exclusive agreement with any company. Right now, developers can use that information for free, which seems like a win-win situation to us… Metro currently reserves the right to withdraw the data in the future, because… the results of the study may show that Metro can earn a substantial amount of money from the intellectual property we produce.[100]

The agency was also concerned about potential liabilities arising from problems with data quality. While local technologists were eager to work with any data at all, even if it was not completely accurate, agency officials were conscious of their responsibility to instill confidence in the transit system, particularly after a tragic train accident on the Red Line in June 2009.

As Emeka Moneme, WMATA's former Chief Administrative Officer, pointed out, "the public assumption is that the data is perfect, clean and accurate. But that's not the case."[101] For example, the agency had three different inventories for the geographic location identifiers of its bus stops, leaving the precise coordinates of some 12,000 stops unclear. Without identifiers for the stops, the GTFS file on Google Transit would be meaningless; people searching online would not know where to wait for the bus. "This is an example of the kind of thing that wasn't an operational problem for WMATA until they decided to pull back the curtain on their data, at which point data quality had to be tightened up."[102] It became an 18-month effort for the agency to consolidate the three bus stop inventories into one. Consultants had to physically go out to thousands of bus stops to log the precise geographic coordinates for each location. Victor Grimes, WMATA's Chief of Enterprise Web Portal and GIS commented that people have "no sense of the timeframe it takes to implement these things, there are 30-year-old problems that have to be fixed."[103]

---

[100] Quoted in Michael Perkins' August 6, 2009 Greater Greater Washington blog post "Metro responds to Google Transit posts." http://greatergreaterwashington.org/post/3145/metro-responds-to-google-transit-posts/

[101] Author interview with Emeka Moneme.

[102] Tom Lee via email communication with the author December 2011.

[103] Author interview with Victor Grimes.

Impressively, the agency did fix these problems on a shoestring budget and well enough to be able to announce a series of data releases in the latter half of 2010. By this point, WMATA had a new General Manager who showed a greater willingness to open up the agency's data. It had also reinstated its contract with Nextbus and were monitoring the service's performance regularly.[104] In August 2010, WMATA provided a public API for real-time Metrorail data. And later that December, it provided the public with access to real-time Metrobus positions. Former Chief Administrative Officer Moneme acknowledged that if had not been for the intense public pressure coming from local bloggers and software developers, "WMATA wouldn't have released any data."[105]

Washington's transit data ecosystem has not flourished as widely as those in Portland, Boston and Chicago, however. While schedules and real time information are publicly available to developers, few applications have been built – just 11 are featured on the agency's Application Gallery as of the writing of this report. Unlike agencies that have benefited from the work of a vibrant community of developers, WMATA lacks any reliable mechanism through which developers can communicate with the agency. A Facebook group was formed at the time of the data releases and provided a conduit of communication between an agency point person, who worked on a contract basis. But since that contractor ceased to be employed by WMATA, the group has been dormant. Kurt Rasche, a DC-based transit enthusiast who keeps a blog on transportation issues, detailed problems with WMATA's elevator and escalator status data feed in an April 2012 blog post. His discussion highlighted the critical role that fostering developer relationships plays in sustaining a community of transparency:

> What's really frustrating about this is the fact that there's just no good way to communicate these types of issues back to WMATA—no dedicated point of contact, no developer mailing list, no developer meetups, nothing.[106]

Even with data in high demand, such as in the case of transit, the absence of developer support from within the agency diminishes the potential for broader public benefits from exposing data and puts in peril the sustainability of such efforts.

---

[104] Washington Metropolitan Area Transit Authority (2010).
[105] Author interview with Emeka Moneme.
[106] Raschke, Kurt (2012).

The Late Adopter: New York's MTA

Metropolitan Transportation Authority

New York's Metropolitan Transportation Authority released operations data to the public later than the other agencies studied in this report. Like in Washington, constraints involved a cautious institutional culture and technical challenges. More so than any other agency examined for this study, a considerable obstacle in releasing operations data for the MTA was the sheer scale of its transit system: the agency's stock includes of 6,000 buses and 468 subway stations serving over eight million riders on an average weekday.[107]

In May 2010 the MTA got onboard with the open data strategy that had already proven fruitful for Portland's TriMet, Boston's MBTA and other transit agencies. The MTA announced the release of multiple datasets at a developers' conference: GTFS schedule and route files for its entire system, weekly turnstile counts, service updates, escalator and elevator status, and performance data.[108] This was a significant victory for local technologists. For months, they had been pressing for the release of MTA data through a coordinated campaign called NYTransitdata.org, led by the non-profit group OpenPlans.[109] They had a clear ask:

> We want bus, subway, and train schedules and route information. The data we want is the same master data that the MTA uses for generating the printed schedules. There's lots of other data that might be useful as well, but schedules and routes could be made available immediately, and would provide the greatest immediate gain.[110]

This group argued that demand existed for this information and that developers were finding ways of getting schedule and route data anyhow – by screen scraping it off the MTA website, filing freedom of information requests or generating their own data. Plus, by this time, smaller agencies around the country had shown that an open data model was a low-cost approach to improving customer information. By keeping its public transit locked up, "New York risk[ed] falling further behind the rest of the country."[111] Making public timely and accurate data that originated officially from the MTA would benefit the agency and developers alike by reducing the risk of applications misrepresenting transit information to customers and saving time for agency managers in fulfilling public data requests.

The MTA's announcement of an open data strategy was a notable turnaround in policy. Less than a year earlier the MTA had sent a cease-and-desist order to a

[107] Figures from the APTA Transit Ridership report, first quarter of 2012. Also, from author's notes of a Harvard Kennedy School presentation by MTA Chairman Jay Walder on March 2, 2011 and MTA Facts and Figures, Subway and Bus ridership http://www.mta.info/nyct/facts/ridership/
[108] MTA Developers Unconference http://www.mta.info/developers/conf.html
[109] See http://nytransitdata.org
[110] http://nytransitdata.org/faq.html#what
[111] http://nytransitdata.org/faq.html#precedent

software developer for providing access to Metro-North commuter rail schedules on his StationStops website and iPhone application.[112] The agency claimed that schedule information was copyrighted intellectual property and expected the developer to pay advance royalty payments and ten percent of future profit from his $2.99 iPhone application. The MTA had a similar claim against another iPhone application using Long Island Railroad schedules. Negative press followed, and by September 2009, the MTA admitted to making a mistake in claiming intellectual property rights over public transit schedules.[113] While the data became freely accessible to the public, the agency continued to charge licensing fees for the use of any MTA symbols, logos or maps on third-party applications.[114] The change in culture came from the top: Jay Walder became the MTA's new chairman and chief executive officer in the fall of 2009. Under Walder's leadership, the MTA pushed forward the disclosure of its operations information.

Still, as of 2010, the agency could not provide real-time data feeds for its network of buses or trains. The main challenge in releasing real-time data to the public reflected a longstanding and unique problem for the MTA. The large scale of the city's bus network combined with the density of Manhattan made GPS tracking a challenge: location signals got lost in the urban canyons of the city. In the meantime, the agency installed countdown clocks on subway station platforms. (And as of May 2012, the MTA was beta-testing the disclosure of subway arrival data to the public in the GTFS-realtime format.) By early 2011, however, the agency made real time bus information available for the B63 bus line in Brooklyn as a pilot project.

The MTA finally solved the bus tracking problem by enhancing existing GPS technology with "dead reckoning", where an algorithm calculates bus locations according to past positions and estimates of speed and elapsed time. In its efforts to implement real-time bus information, the MTA also innovated by building its own its own cost-effective, open source bus tracking system without relying on vendors. Existing, proprietary vendor contracts for real-time bus information were too costly to implement for a system as large as New York's. Instead, the agency achieved economies of scale by developing BusTime, the MTA's in-house hardware and software solution for tracking bus locations.[115] For on-bus hardware, the MTA incorporated off-the-shelf, enhanced-GPS components by piggybacking on the installation of hardware for a new payment system. The BusTime system software was adapted and customized from the OneBusAway open source code originally built for Seattle by a student at the University of Washington. "The result is a system that delivers great results to our customers while being highly cost-effective, fast to deploy, simple to maintain and operate,

[112] See Dana Oshiro's August 20, 2009 account "NY Transportation Authority Cites Schedules as Copyrighted Material."
[113] See Heather Haddon's August 24, 2009 article, "MTA looks for a cut from mobile phone applications." A month later Michael Grynbaum writes in the New York Times, "MTA is Easing Its Strict, Sometimes Combative, Approach to Outside Web Developers."
[114] See the MTA policy on licensing logos, symbols and maps at http://www.mta.info/developers/
[115] For a full explanation of the technologies behind MTA's BusTime see http://bustime.mta.info/wiki/Technology

and supports expansions in the future as time and money allow… First and foremost, the MTA BusTime system is open to external developers to create their own real-time bus applications using the Bus Time data feed."[116] BusTime was piloted along Brooklyn's B63 bus line in January 2011, and a year later, the MTA had implemented BusTime on all 800 buses in Staten Island.[117]

By embracing an open data strategy after other agencies had done so, the MTA was able to build upon prior experiences and forge its own path in innovating a low-cost, open system for delivering real-time bus information to developers and customers. In that sense, the MTA shifted from being a "late adopter" in the diffusion of innovation curve for transit data disclosure, to joining Portland's TriMet as an innovator of open source systems for generating transit data.

In the two years since the MTA released its first datasets, independent software developers have built nearly 70 applications. These are the product of several rounds of apps contests held by the MTA and the city of New York (Big Apps contests). Transit applications are always popular in apps contests since the use of bus and rail data is well suited to mobile platforms with geo-locational capabilities like smartphones. Many of these applications are also anticipating the citywide rollout of real-time bus and subway information. Given that New York City has the greatest proportion of transit riders in the United States, with over 55 percent of residents using public transit to commute to work,[118] there is great potential for independent developers to meet substantial market demand for real-time transit information.

[116] MTA BusTime, http://bustime.mta.info/wiki/Technology
[117] Kabak, Benjamin, 2012.
[118] Data from the 2010 American Community Survey, cited in http://en.wikipedia.org/wiki/List_of_U.S._cities_with_high_transit_ridership

Appendix IV: List of Interviews

Chicago
Chris Cieslak, *software developer, Buster application*
Tony Coppoletta, *Manager, External Electronic Communications, Chicago Transit Authority*
Jakob Eriksson, *University of Illinois at Chicago BITS Lab, TransitGenie application*
Graham Garfield, *General Manager of Customer Information, Chicago Transit Authority*
Daniel X. O'Neil, *Executive Director, Smart Chicago Collaborative*
Jason Shah, *software developer, Chicago Transit Tracker*
Piyushimita Thakuriah (Vonu), *Associate Professor, University of Illinois at Chicago*
John Tolva, *Chief Technology Officer, City of Chicago*

Boston
David Barker, *Manager of Operations Technology, Massachusetts Bay Transportation Authority*
Dan Choi, software developer, *Kaja Software and OpenMBTA application*
Chris Dempsey, *former Deputy Chief of Staff, Massachusetts Department of Transportation*
Jared Egan, *software developer, CatchTheBus and CatchTheT*
Josh Robin, *Director of Innovation, Massachusetts Bay Transportation Authority*
Laurel Ruma, *former Gov 2.0 Evangelist, O'Reilly Media*
George Schneeloch, *software developer, Boston Bus Maps*

Washington DC
David Alpert, *Editor-in-Chief and Founder, Greater Greater Washington blog*
Victor Grimes, *Chief of Enterprise Web Portal & GIS, Department of Information Technology, Washington Metropolitan Area Transit Authority*
Tom Lee, *Director, Sunlight Labs*
Emeka Moneme, *Executive Managing Director, Carmen Group*
Luke Rosiak, *former Local News Deputy Editor for Transportation & Development, The Washington Post*
Kurt Raschke, *transportation blogger, Raschke on Transport*

New York
Fredericka Cuenca, *former Chief of Staff, Metropolitan Transportation Authority*
Mike Frumin, *Systems Engineering Manager, Metropolitan Transportation Authority*
Nick Grossman, *Co-founder and former Managing Director, Civic Commons*
Peter Harris, *Director of Market Research, Metropolitan Transportation Authority*
Frank Hebbert, *Director of Civic Works, OpenPlans*
Sarah Kaufman, *former Projects Coordinator for Best Practices and Strategic Improvements, New York City Transit*
Sunil Nair, *Senior Director, Bus Customer Information Systems, Metropolitan Transportation Authority*

Portland, OR
John McBride, *software developer, myTrimet.com application*
Bibiana McHugh, *IT Manager of Geographic Information Systems, TriMet*
Max Ogden, *software developer, PDXAPI*
Chris Smith, *transit advocate with Portland Transport & developer, TransitSurfer*
Andrew Wallace, *software developer, PDXBus application*

Appendix V: Surveys

MTA B63 Pre-wave Survey

Conducted by Francisca Rojas and Candace Brakewood in coordination with the MTA along the B63 bus route in Brooklyn, New York in January 2011. This pre-wave survey was conducted prior to the availability of real-time bus arrival information as part of the MTA's BusTime pilot project.

2011

B63 LOCAL BUS SERVICE

Customer Travel Survey

SI USTED QUIERE UNA VERSIÓN EN ESPAÑOL DE ESTA ENCUESTA, POR FAVOR PREGUNTE AL REPRESENTANTE QUE LE DIO ESTE DOCUMENTO POR LA VERSIÓN EN ESPAÑOL. ¡MUCHAS GRACIAS!

Dear MTA Bus Customer:

Please take some time to complete this survey being funded in part by the US Department of Transportation and administered by researchers affiliated with Harvard University and the Massachusetts Institute of Technology (MIT) in cooperation with the MTA. Your input will help provide information that will help the MTA improve local bus service on this bus route.

Earn a chance to win a $50 MetroCard by completing this survey. We will award prizes to 10 customers who fully complete the survey. Winners will be selected at random.

When you have completed the survey, please return it to our representative on your bus today.

Thanks for completing this survey!

MTA B63 Pre-wave Survey

ABOUT YOUR TRIP

1 In the past seven (7) days, how many one-way trips did you take on this bus route? A round-trip counts as two one-way trips.

________ (Write number of trips)

2 At what bus stop did you board this bus today?

________ (Intersection, or cross streets, nearest your bus stop)

3 What time of day did you board this bus today?

________ (Time you boarded bus today) AM PM (circle one)

4 Did you transfer from any of the following forms of transportation before boarding this bus today?

(CIRCLE ALL THAT APPLY)

Local bus	1	Please specify route # ________
Subway	2	Please specify subway line ________
Long Island Rail Road	3	Please specify branch ________
Other	4	Please specify ________

5 Where will you exit this bus?

________ (Intersection, or cross streets, nearest your bus stop)

6 When you get off this bus, will you be transferring to:

(CIRCLE ALL THAT APPLY)

Local bus	1	Please specify route # ________
Subway	2	Please specify subway line ________
Long Island Rail Road	3	Please specify branch ________
Other	4	Please specify ________

7 Which direction is this bus headed?

(CIRCLE ONLY ONE NUMBER)

Headed toward downtown Brooklyn	1
Headed away from downtown Brooklyn	2

8 How many minutes did you wait for this bus to arrive today?

________ (minutes)

9 How many minutes do you usually wait for this bus to arrive?

________ (minutes)

10 What is the primary purpose of this bus trip?

(CIRCLE ONLY ONE NUMBER)

Work/work-related	1
Leisure/recreation	2
Personal business (doctor)	3
Shopping	4
School	5
Other	6

11 Before you boarded the bus for this trip today, did you do any of the following?

	Yes	No
Showed up without checking the bus schedule or service status	1	2
Checked the schedule posted at the bus stop	1	2
Checked the schedule on mta.info	1	2
Checked the schedule on another website (i.e. Google Transit)	1	2
Checked online if buses were running on time	1	2
Received a text message service status alert from MTA	1	2
Walked to the next bus stop to catch the bus sooner	1	2
Started walking toward your destination	1	2

12 On this trip, did you spend all your wait time at the bus stop or did you spend some of it elsewhere?

Spent all your time at the bus stop	1	GO TO Q.14
Spent some of it elsewhere	2	GO TO Q.13

13 Where else did you spend time while waiting at the bus stop?

Store	1	
Coffee shop	2	
Home	3	
Other	4	Please specify where ________

14. Was the bus at the bus stop when you arrived or close enough that you could see it coming?

At bus stop or close enough to see it coming	1	SKIP TO Q16
Not close enough to see	2	ASK Q15

15 How did you feel about waiting for *this* bus today? Please indicate below if you agree or disagree with each of the following by circling one of the numbers below:

	Agree Strongly	Agree Somewhat	Disagree Somewhat	Disagree Strongly
You felt confident the bus would arrive soon, based on your experience	1	2	3	4
You felt confident the bus would arrive soon, based on available information	1	2	3	4
You felt frustrated not knowing when the bus would arrive	1	2	3	4
You felt safe waiting at the bus stop	1	2	3	4

16 How often do you do each of the following to obtain bus schedule information? For each item, please CIRCLE one of the numbers below.

	Very Often	Somewhat Often	Rarely	Never
Just show up at bus stop without checking the schedule or service status	1	2	3	4
Read schedules posted at the bus stop	1	2	3	4
Go online to mta.info to get the schedule	1	2	3	4
Go online to mta.info to get service status for this bus route	1	2	3	4
Go to a non-MTA website for schedules or service status	1	2	3	4
Walk to another bus stop on this route so you catch the bus sooner	1	2	3	4
Give up waiting for this bus and walk to another local bus route, a subway or express bus?	1	2	3	4
Wait at the bus stop but use the time to make phone calls, go online, shop, etc.	1	2	3	4

17 Please tell us how satisfied you are with various aspects of local bus service. For each aspect of service, please CIRCLE one of the numbers below to indicate your satisfaction with that aspect of local bus service.

	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
Overall local bus service on this route	1	2	3	4
Frequency of service	1	2	3	4
Predictability of travel time	1	2	3	4
Availability of schedules at bus stop	1	2	3	4
Availability of bus schedules online/Internet	1	2	3	4
How long you have to wait for a bus	1	2	3	4
Knowing how far away the next bus is	1	2	3	4
Overall cleanliness inside the bus	1	2	3	4
Temperature on bus	1	2	3	4

MTA B63 Pre-wave Survey

18 Do you own or have access to a car when you need one?

Yes	1
No	2

19 Which of the following have you used in the past 30 days?

	Fill in Circles if Used in Past 30 Days
Internet on computer	❍
Internet on phone	❍
Cell phone	❍
Text messaging	❍
Your own email address	❍
Smart phone (Please specify type below)	❍
iphone	❍
Android	❍
Blackberry	❍
Other	❍

The following questions are for classification purposes only. The information helps us ensure that all groups are represented in the survey.

D1 What is your home zip code? ______________________

D2 Are you ...?

Male 1 Female 2

D3 In which of the following categories does your age fall?

16 or 17	1	***RESPONDENT MUST BE AGE 16 OR OLDER***
18 to 29	2	
30 to 39	3	
40 to 54	4	
55 to 64	5	
65 and over	6	

D4 Are you Hispanic? Yes 1 No 2

D5 Are you ...?

White	1
African-American or Black	2
American Indian or Alaska Native	3
Asian	4
Other	5 Please specify: ____________

D6 To ensure our study is representative of all income groups in the city, please indicate which of the following categories best describes you total household income for 2010?

Below $15,000	1
$15,000 to less than $35,000	2
$35,000 to less than $50,000	3
$50,000 to less than $75,000	4
$75,000 to less than $100,000	5
$100,000 or more	6

D7 Are you ...?

Employed full-time	1
Employed part-time	2
Not employed	3

D8 What is the last grade of school completed?

Less than high school	1	Some college	4
High school/GED	2	4-year college	5
Technical/vocational school	3	Graduate school	6

To be eligible for the MetroCard lottery, we need your phone number or email address. Harvard University and the Massachusetts Institute of Technology (MIT) will keep your contact information in complete confidence. We guarantee that your responses will not be identified with you personally.

Telephone number (________) _______ - ______________

email address ____________________ **@** ______________

D9 Please check this box if you want to be contacted for future MTA research. ☐

THANK YOU!

MTA B63 Post-wave Survey

Conducted by the Lieberman Research Group on behalf of the MTA and in consultation with Francisca Rojas and Candace Brakewood between June 10 and 14, 2011. The survey was distributed to riders on the B63 bus route, for which BusTime became available February 1, 2011, and to riders on the B62 bus route as a control.

2011

MTA Customer Travel Survey

B63 LOCAL BUS SERVICE

SI USTED QUIERE UNA VERSIÓN EN ESPAÑOL DE ESTA ENCUESTA, POR FAVOR PREGUNTE AL REPRESENTANTE QUE LE DIO ESTE DOCUMENTO POR LA VERSIÓN EN ESPAÑOL. ¡MUCHAS GRACIAS!

Dear B63 Bus Customer:

Please take some time to complete this survey. It includes questions about your travel on the B63 and about new MTA bus information initiatives.

IF YOU COMPLETE THE FULL SURVEY, YOU WILL BE ENTERED INTO A DRAWING THAT WILL AWARD $20 METROCARDS TO 25 LOCAL BUS CUSTOMERS.

Employees and families of the MTA and its constituent agencies and firms administering the survey are not eligible for the prize drawing.

Please return the completed survey to our representative on your bus today or return it by mail - no postage necessary.

Thanks for completing this survey!

MTA B63 Post-wave Survey

1 In the past seven (7) days, how many one-way trips did you take on this bus route including today's trip?

_______________ (Write number of trips)

2 What time of day did you board this bus today? ___________ **AM PM** (circle one)

3 Where did you board this bus today?

_______________ (Intersection, or cross streets, nearest your bus stop)

4 Where will you exit this bus?

_______________ (Intersection, or cross streets, nearest your bus stop)

5 Which direction is this bus headed? (circle one number)

Toward Cobble Hill (downtown Brooklyn)	1
Toward Ft. Hamilton (away from downtown Brooklyn)	2

6 Where was the bus when you arrived at the bus stop today? (circle one number)

At bus stop or close enough to see it coming	1	**Skip to Q10**
Not close enough to see the bus	2	**Go to Q7**

7 Total minutes you waited for this bus to arrive today ______(minutes)

8 Total minutes you usually wait for this bus to arrive ______(minutes)

9 How did you feel about waiting for this bus today? Please indicate below if you agree or disagree with each statement below by circling one of the numbers below for each item:

	Disagree Strongly	Disagree	Agree	Agree Strongly
I felt confident the bus would arrive soon, based on my experience	1	2	3	4
I felt confident the bus would arrive soon, based on available information	1	2	3	4
I felt frustrated not knowing when the bus would arrive	1	2	3	4

10 What is the primary purpose of this local bus trip? _______________

11 Please tell us how satisfied you are with local bus service on this route.

	Very Dissatisfied	Dissatisfied	Satisfied	Very Satisfied	N/A
Overall local bus service on this route	1	2	3	4	9
Frequency of service	1	2	3	4	9
Reliability of service	1	2	3	4	9
Availability of schedules at bus stop	1	2	3	4	9
How long you have to wait for the bus	1	2	3	4	9
Knowing how far away the next bus is	1	2	3	4	9

12 Have you heard or read about MTA BusTime, a new way for B63 customers to get information on how far away the next bus is from their bus stop?

Circle Your Answer Below

Yes	1	**Go to Q13**
No	2	**Skip to Q22**

13 Where did you hear or read about MTA BusTime?

Circle All That Apply to You.

MTA BusTime marketing material	1	On MTA website	4
BusTime info at bus stop	2	News article or word of mouth	5
BusTime info displayed in a storefront window	3	Other (please describe) _______	

14 IF YOU HAVE NEVER USED MTA BusTime, which of the following best explains why not?

Circle All That Apply to You

Don't need it	1
It's hard to understand how to use MTA BusTime	2
Tried it but could not figure it out	3
I do not have necessary technology (phone, computer, etc) to get BusTime information	4
Other reasons (specify) _______	5

IF YOU HAVE NEVER USED MTA BUSTIME, Skip to Q22

IF YOU HAVE USED MTA BUSTIME

15 MTA customers can get BusTime information several ways. Which have you tried? PLEASE ANSWER EACH QUESTION.

	Successfully Got BusTime Info This Way	Tried But Could Not Get BusTime Info This Way	Never Tried to Get BusTime Info This Way
BusTime on a desktop computer	1	2	3
Text messaged BusTime	1	2	3
BusTime on mobile website using a smartphone (iPhone, Android, etc.)	1	2	3

15b Have you seen BusTime info in a storefront window? Yes 1 No 2

16 How satisfied are you with MTA BusTime overall? ***Please rate only the methods you have used to obtain MTA BusTime information:***

Circle Your Answer Below

Rate methods used	Very Dissatisfied	Dissatisfied	Satisfied	Very Satisfied	N/A
From storefront window	1	2	3	4	9
On desktop computer	1	2	3	4	9
By text messaging	1	2	3	4	9
MTA's mobile web site	1	2	3	4	9

17 How do you usually get BusTime information?

Check Only One Box

Storefront window	☐
Desktop computer via Internet	☐
Text messaging BusTime on a mobile/cell phone	☐
MTA's mobile web site, using a smartphone (iPhone, Android, etc.)	☐
Other (specify method):_______________	☐

18 Is it easy or difficult to:

	Very Easy	Easy	Difficult	Very Difficult	N/A
Find instructions on how to use BusTime?	1	2	3	4	9
Follow instructions for using BusTime?	1	2	3	4	9
Get BusTime information?	1	2	3	4	9

19 How often do you use MTA BusTime?

Circle Your Answer Below

Most of my B63 trips	1
About half of my B63 trips	2
Occasionally	3
Rarely	4

20 How useful is BusTime information in estimating how long you have to wait until the next bus arrives?

Circle Your Answer Below

Very useful	1
Useful	2
Not too useful	3
Not at all useful	4

21 Would you like to see MTA BusTime available for all bus routes?

Circle Your Answer Below

Yes	1	
No	2	Why not? _______________

ALL CUSTOMERS PLEASE ANSWER Q22

22 On this trip, did you spend all your wait time at the bus stop or did you spend some of it elsewhere?

Circle Your Answer Below

Spent some of it elsewhere	1	**Go to Q23**
Spent all your time at bus stop	2	**Skip to Q24**

23 Where else did you spend time while waiting for the bus?

Circle Your Answer Below

Store	1	Coffee shop	3
Home or work	2	Other	4 Please specify where _______

ALL CUSTOMERS PLEASE ANSWER ALL REMAINING QUESTIONS

24 How often have you done each of the following on this bus route?

	Very Often	Somewhat Often	Rarely	Never
Show up at bus stop without checking bus schedule or service status	1	2	3	4
Check bus schedule posted at bus stop	1	2	3	4
Check bus schedule or service status				
On MTA website	1	2	3	4
On another website	1	2	3	4
On a transit app	1	2	3	4
Walk to next bus stop to catch bus sooner	1	2	3	4
Start walking toward destination	1	2	3	4
Give up waiting for bus and walk to another local bus route, subway, or express bus	1	2	3	4
Use wait time to shop or do something else, then walk to bus stop in time to catch bus	1	2	3	4

MTA B63 Post-wave Survey

25 Which of the following have you used in the past 30 days?
(circle the number of each item used in past 30 days)

Cell phone	1	Internet on phone	3	Personal email	5
Text messaging	2	Internet on computer	4	None of the above	6

The following questions are for classification purposes only. The information helps us ensure that all groups are represented in the survey.

D1 Are you …? Male 1 Female 2

D2 What is your age?

16 or 17	1	***RESPONDENT MUST BE AGE 16 OR OLDER***
18 to 29	2	
30 to 39	3	
40 to 54	4	
55 to 64	5	
65 and over	6	

D3 Are you Hispanic? Yes 1 No 2

D4 Are you …?

White	1	Racially mixed	4
African-American or Black	2	Other	5
Asian	3		

D5 To ensure our study is representative of all income groups in the city, please tell us which of the following categories best describes your total household income?

Below $15,000	1	$50,000 to less than $75,000	4
$15,000 to less than $35,000	2	$75,000 to less than $100,000	5
$35,000 to less than $50,000	3	$100,000 or more	6

D6 What is the last grade of school completed?

Less than high school	1	Some college	4
High school/GED	2	4-year college	5
Technical/vocational school	3	Graduate school	6

To be eligible for the MetroCard drawing, we need your phone number or email address. The MTA and its survey consultant will keep your contact information confidential

Telephone number (__________) ________ - ________________

email address ________________________ @ ______________

Check box ☐ if OK for the MTA to contact you with additional questions.

MTA B62 Control Route Survey

This was the control route survey along the B62 route, distributed in concert with the B63 BusTime survey in June 2011.

2011
MTA Customer Travel Survey
B62 LOCAL BUS SERVICE

SI USTED QUIERE UNA VERSIÓN EN ESPAÑOL DE ESTA ENCUESTA, POR FAVOR PREGUNTE AL REPRESENTANTE QUE LE DIO ESTE DOCUMENTO POR LA VERSIÓN EN ESPAÑOL. ¡MUCHAS GRACIAS!

Dear B62 Bus Customer:

Please take some time to complete this survey being conducted by the MTA. It includes questions about your travel on the B62.

IF YOU COMPLETE THE FULL SURVEY, YOU WILL BE ENTERED INTO A DRAWING THAT WILL AWARD $20 METROCARDS TO 25 LOCAL BUS CUSTOMERS. *Employees and families of the MTA and its constituent agencies and firms administering the survey are not eligible for the prize drawing.*

When you have completed the survey, please return it to our representative on your bus today or return it by mail - no postage necessary.

Thanks for completing this survey!

MTA B62 Control Route Survey

1 In the past seven (7) days, how many one-way trips did you take on this bus route including today's trip?

______________ (Write number of trips)

2 What time of day did you board this bus today? __________ **AM PM** (circle one)

3 Where did you board this bus today?

____
(Intersection, or cross streets, nearest your bus stop)

4 Where will you exit this bus?

____
(Intersection, or cross streets, nearest your bus stop)

5 Which direction is this bus headed? (circle one number)

Toward Downtown Brooklyn	1
Toward Queens Plaza	2

6 Where was the bus when you arrived at the bus stop today? (circle one number)

At bus stop or close enough to see it coming	1	**Skip to Q10**
Not close enough to see the bus	2	**Go to Q7**

7 Total minutes you waited for this bus to arrive today ______(minutes)

8 Total minutes you usually wait for this bus to arrive ______(minutes)

9 How did you feel about waiting for this bus today? Please indicate below if you agree or disagree with each statement below by circling one of the numbers below for each item:

	Disagree Strongly	Disagree	Agree	Agree Strongly
I felt confident the bus would arrive soon, based on my experience	1	2	3	4
I felt confident the bus would arrive soon, based on available information	1	2	3	4
I felt frustrated not knowing when the bus would arrive	1	2	3	4

10 What is the primary purpose of this local bus trip? ______________

11 On this trip, did you spend all your wait time at the bus stop or did you spend some of it elsewhere?

Circle Your Answer Below

Spent some of it elsewhere	1	**Go to Q12**
Spent all your time at bus stop	2	**Skip to Q13**

12 Where else did you spend time while waiting for the bus?

Circle Your Answer Below

Store	1	Coffee shop	3
Home or work	2	Other	4 Please specify where ___________

ALL CUSTOMERS PLEASE ANSWER ALL REMAINING QUESTIONS

13 How often have you done each of the following on this bus route in the past 3 months? *CIRCLE THE NUMBER below that that best applies to you for each item.*

	Very Often	Somewhat Often	Rarely	Never
Show up at bus stop without checking bus schedule or service status	1	2	3	4
Check bus schedule posted at bus stop	1	2	3	4
Check bus schedule or service status				
On MTA website	1	2	3	4
On another website	1	2	3	4
On a transit app	1	2	3	4
Walk to next bus stop to catch the bus sooner	1	2	3	4
Start walking toward destination	1	2	3	4
Give up waiting for bus and walk to another local bus route, subway, or express bus	1	2	3	4
Use wait time to shop or do something else, then walk to bus stop in time to catch bus	1	2	3	4

14 Please tell us how satisfied you are with local bus service on this bus route.

	Very Dissatisfied	Dissatisfied	Satisfied	Very Satisfied	N/A
Overall local bus service on this route	1	2	3	4	9
Frequency of service	1	2	3	4	9
Reliability of service	1	2	3	4	9
Availability of schedules at bus stop	1	2	3	4	9
How long you have to wait for the bus	1	2	3	4	9
Knowing how far away the next bus is	1	2	3	4	9

15 Which of the following have you used in the past 30 days?

(circle the number of each item used in past 30 days)

Cell phone	1	Internet on phone	3	Personal email	5
Text messaging	2	Internet on computer	4	None of the above	6

The following questions are for classification purposes only. The information helps us ensure that all groups are represented in the survey.

D1 Are you ...? Male 1 Female 2

D2 What is your age?

16 or 17	1	***RESPONDENT MUST BE AGE 16 OR OLDER***
18 to 29	2	
30 to 39	3	
40 to 54	4	
55 to 64	5	
65 and over	6	

D3 Are you Hispanic? Yes 1 No 2

D4 Are you ...?

White	1	Racially mixed	4
African-American or Black	2	Other	5
Asian	3		

D5 To ensure our study is representative of all income groups in the city, please tell us which of the following categories best describes your total household income?

Below $15,000	1	$50,000 to less than $75,000	4
$15,000 to less than $35,000	2	$75,000 to less than $100,000	5
$35,000 to less than $50,000	3	$100,000 or more	6

D6 What is the last grade of school completed?

Less than high school	1	Some college	4
High school/GED	2	4-year college	5
Technical/vocational school	3	Graduate school	6

To be eligible for the MetroCard drawing, we need your phone number or email address. The MTA and its survey consultant will keep your contact information confidential

Telephone number (________**)** ________ **-** ______________

email address ____________________ **@** ____________

Check box ☐ **if OK for the MTA to contact you with additional questions.**

MBTA Commuter Rail Pre-wave Survey

Conducted by Francisca Rojas and Candace Brakewood in coordination with the MBTA on the Newbury/Rockport and Worcester lines of the Commuter Rail system in May 2011. This pre-wave survey was conducted prior to the availability of real-time arrival information on the MBTA's Commuter Rail system.

1. TIME: _____ AM / PM (Circle) **2. LINE**: O Worcester O Newburyport O Rockport **3. DIRECTION**: O Inbound O Outbound

--PLEASE DO NOT WRITE ABOVE THIS LINE ---

Dear MBTA Commuter Rail Customer:

Please take some time to complete this survey being administered by researchers affiliated with Harvard University and the Massachusetts Institute of Technology (MIT) in cooperation with the MBTA. Your input will provide information that will help the MBTA improve Commuter Rail service, and researchers will use the results of the survey to better understand how transit riders get information and make transportation decisions.

Please read the information below and ask questions about anything you do not understand before deciding whether or not to participate in this survey.

- The survey should take you about 5-8 minutes to complete.
- This survey is completely voluntary. You have the right not to answer any question and to stop the survey at any time.
- We do not know of any risks to you if you decide to participate in this survey.
- We guarantee that your responses will not be identified with you personally. Your answers will remain completely confidential and will only be used for research and evaluation purposes.

If you have any questions or concerns about completing the questionnaire or about being in this study, you may contact Francisca Rojas at francisca_rojas@hks.harvard.edu or Candace Brakewood at ceb@mit.edu. If you have any concerns about your rights as a participant in this study you may contact the Harvard IRB via email at jcalhoun@fas.harvard.edu or by telephone (617-495-5459) or the MIT COUHES at jadams@mit.edu or by telephone (617-253-6787)

IF YOU COMPLETE THE FULL SURVEY, YOU WILL BE ENTERED INTO A LOTTERY THAT WILL AWARD CHARLIE CARDS TO 10 COMMUTER RAIL CUSTOMERS.

To be eligible for the Charlie Card lottery, we need your phone number or email address. Your contact information will be maintained by the researchers in a completely confidential data file.

Telephone number:

____

Email address:

______________________________@______________________

❑ Please check this box if you want to be contacted for future MBTA research.

When you have completed the survey, please return it to our representative on your train today.

THANKS FOR COMPLETING THIS SURVEY!

MBTA Commuter Rail Pre-wave Survey

1. How many days a week do you ride this rail line?

O Less than 1 day **O** 3 days **O** 6 days
O 1 day **O** 4 days **O** 7 days
O 2 days **O** 5 days **O** I'm only visiting Boston

2. At what Commuter Rail station did you **board** this train today? ___________________________
(station name, e.g. North Station)

3. At what Commuter Rail station will you **exit** the train today? ______________________________
(station name, e.g. North Station)

4. How did you get to the station where you boarded this train (reported in question 2)?

O Drove or rode in a personal vehicle and parked at or near the station **O** Private shuttle van/bus
O Dropped off by personal vehicle that did not park **O** Taxi **O** THE RIDE
O Walked directly (from work, school, home, etc.) **O** Bicycle **O** MBTA Boat
O MBTA Red, Blue, Orange or Green line subway train **O** MBTA Bus or Silver Line
O Other _____________

5. Where will this **one-way** trip end?

O At work **O** At a doctor or other personal business
O At school **O** At a work-related errand or meeting
O At home **O** At a restaurant or social or recreational activity
O At a store **O** Other _____________

6. How time-sensitive is **this** trip?

O I cannot be late **O** I can be a few minutes late **O** I have flexibility

7. How will you get to your **destination** from the station reported in question 3?

O Drive or ride in a personal vehicle and park at or near the station **O** Private shuttle van/bus
O Pick up by a personal vehicle that will not park **O** Taxi **O** THE RIDE
O Walk directly from the station **O** Bicycle **O** MBTA Boat
O MBTA Red, Blue, Orange or Green line subway train **O** MBTA Bus or Silver Line
O Other _____________

8. How long did you wait at the commuter rail station **today**? _________________________
(Write number, e.g. 7 minutes)

9. How long do you **usually** wait at the commuter rail station? _______________________
(Write number, e.g. 7 minutes)

10. While you were waiting for the train today, did you do any of the following? **Check all that apply.**

❑ Spent all of your wait time at the train station or in a car at the station
❑ Used the time waiting for the train to make phone calls, listen to music, go online, etc.
❑ Spent some time at a nearby store ❑ Spent some time at a coffee shop or restaurant
❑ Spent some time at home ❑ Spent some time at work ❑ Other:________________

11. Which devices/technologies have you used in the past 30 days? **Check all that apply.**

❑ Blackberry ❑ iPhone ❑ Android Smartphone ❑ Other Smartphone
❑ Laptop Computer ❑ Desktop Computer ❑ Tablet Computer (iPad)
❑ Cell Phone ❑ Text Messaging ❑ None of the Above

MBTA Commuter Rail Pre-wave Survey

12. How often do you do each of the following before boarding **Commuter Rail** trains?

	Never	**Sometimes** (A few rides)	**Often** (Most rides)	**Always** (Every ride)
a. Show up at the station *without* checking the schedule or service status	1	2	3	4
b. Read a *paper schedule* before departing for the station	1	2	3	4
c. Go to *mbta.com* to:				
1. check commuter rail schedules	1	2	3	4
2. use the trip planner	1	2	3	4
3. check service alerts and advisories	1	2	3	4
d. Go to a *non-MBTA website* for trip planner or schedules (e.g. Google Transit)	1	2	3	4
e. Receive a T-alert via *email or text message (SMS)*	1	2	3	4
f. Consult the commuter rail *schedule posted at the train station*	1	2	3	4
g. Consult the commuter rail *system map posted at the train station*	1	2	3	4
h. Listen to *MBTA radio (1630 AM)* in my car while waiting for the train to arrive	1	2	3	4
i. Watch the train arrival *announcement scrolling on the LED sign* at the station	1	2	3	4
j. Watch the *flashing light* in the station parking lot indicating my train's arrival	1	2	3	4

13. Several measures of service quality are listed below. Please circle a number after each measure to indicate your opinion on the **level of service** you typically experience on MBTA Commuter Rail trips and its **level of importance** to you. (Circle NA for any measures that don't apply).

	Level of Importance					**Level of Service Quality**					
	Not Important		Neutral		Very Important	Poor		Average		Excellent	Not Applicable
a. On-time performance (reliability)	1	2	3	4	5	1	2	3	4	5	NA
b. How long you wait for the train	1	2	3	4	5	1	2	3	4	5	NA
c. Amount of time between trains	1	2	3	4	5	1	2	3	4	5	NA
d. Arriving at your destination on time	1	2	3	4	5	1	2	3	4	5	NA
e. Personal safety at the station	1	2	3	4	5	1	2	3	4	5	NA
f. Availability of shelters/seats at the station	1	2	3	4	5	1	2	3	4	5	NA
g. Availability of parking at the station	1	2	3	4	5	1	2	3	4	5	NA
h. Amenities at stations (dining/food facilities, restrooms)	1	2	3	4	5	1	2	3	4	5	NA
i. Courtesy of train crews	1	2	3	4	5	1	2	3	4	5	NA
j. Clear & timely announcements of approaching stations	1	2	3	4	5	1	2	3	4	5	NA
k. Availability of schedule information	1	2	3	4	5	1	2	3	4	5	NA
l. Availability of maps and route information	1	2	3	4	5	1	2	3	4	5	NA
m. Ease of getting information via the internet	1	2	3	4	5	1	2	3	4	5	NA
n. Effectiveness of T-Alerts for incidents	1	2	3	4	5	1	2	3	4	5	NA
o. Explaining reasons for delays or other problems	1	2	3	4	5	1	2	3	4	5	NA
p. Overall commuter rail service on this line	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**NA**

MBTA Commuter Rail Pre-wave Survey

Last year, the MBTA made available "real time" information about the precise GPS location and predicted arrival times for buses and subway trains on the T. (This is not the same as scheduled arrival times). Riders can access this real time information through websites, on cell phones (via text message/SMS) and on smartphones (by downloading transit "apps" for the T made by third-party software developers).

14. Were you aware of the availability of real time information for MBTA buses and subway trains?
    O Yes, I was aware of real-time bus and subway train information for the T.
    O No, I was not aware of real-time information (skip the next question and go to question D1)

15. When riding MBTA bus or subway trains in Boston, how do you consult real time location or arrival information? For which modes of transportation? **Check all that apply.**

	MBTA buses	MBTA subway	Not Applicable
On a desktop or laptop computer	❑	❑	❑
On my smartphone through "apps"	❑	❑	❑
On my smartphone through a mobile website/web browser	❑	❑	❑
On my cell phone through SMS/text messaging	❑	❑	❑

The following questions are for classification purposes only. This information helps the MBTA Commuter Rail provide the best service to all our customers.

D1. What is your home zip code? __________

D2. What is your gender?
O Male **O** Female

D3. What is your age?
O 18 or under **O** 25-34 **O** 45-64
O 19-24 **O** 35-44 **O** 65 and over

D4. How do you self-identify by race? (Check all that apply.)
O American Indian or Alaska Native **O** Asian
O Black or African American **O** White
O Native Hawaiian or Pacific Islander **O** Other________

D5. Are you Hispanic/Latino?
O Yes **O** No

D6.What is your annual combined household income?
O Under $20,000 **O** $40,000-$49,999 **O** $75,000-$99,999
O $20,000-$29,999 **O** $50,000-$59,999 **O** $100,000 or more
O $30,000-$39,999 **O** $60,000-74,999